Published: 16 June 2025

Citation in BibTeX format

ICAIL 2025: 20th International Conference on Artificial Intelligence and Law
June 16 - 20, 2025
IL, Chicago, USA

RESEARCH-ARTICLE

Tipping the Scales: Balancing Principles

MICHAŁ ARASZKIEWICZ, Jagiellonian University, Krakow, Małopolska, Poland

TOMASZ ZUREK, University of Amsterdam, Amsterdam, Noord-Holland, Netherlands

ADAM ZACHARY WYNER, Swansea University, Swansea, West Glamorgan, Wales, U.K.

TOM M VAN ENGERS, University of Amsterdam, Amsterdam, Noord-Holland, Netherlands

Open Access Support provided by:

Jagiellonian University

University of Amsterdam

Swansea University

ICAIL '25: Proceedings of the Twentieth International Conference on Artificial Intelligence and Law (June 2025)
https://doi.org/10.1145/3769126.3769208
ISBN: 9798400719394

Tipping the Scales: Balancing Principles

MichałAraszkiewicz*
Department of Legal Theory, Faculty of Law and Administration
Jagiellonian University
Krakow, Poland
michal.araszkiewicz@uj.edu.pl

Tomasz Zurek
Complex Cyber Infrastructure, Informatics Institute
University of Amsterdam
Amsterdam, Netherlands
t.a.zurek@uva.nl

Adam Wyner
Department of Computer Science
Swansea University
Swansea, United Kingdom
a.z.wyner@swansea.ac.uk

Tom van Engers
Leibniz Institute
University of Amsterdam & TNO
Amsterdam, Netherlands
t.m.vanengers@uva.nl

Abstract

The paper addresses the problem of making legal decisions about regulations such that they are compliant with legal principles. While decision-making using legal rules is well-developed in the literature, few researchers address reasoning with principles. A key difference between applying rules versus principles is that rules have a binary character, they are applied or not, while principles can be applied to a certain extent and in the case of conflicting principles, they require balancing instead of defeat. This paper provides a novel approach to balancing principles and deciding amongst regulations by proportional optimisation of values associated with regulations and principles. The approach can be generalised to other areas of decision-making.

CCS Concepts

• **Computing methodologies → Reasoning about belief and knowledge**; • **Applied computing → Law**.

Keywords

legal compliant decisions, balancing, decision making, principles, computational model

ACM Reference Format:
MichałAraszkiewicz, Tomasz Zurek, Adam Wyner, and Tom van Engers. 2025. Tipping the Scales: Balancing Principles. In *20th International Conference on Artificial Intelligence and Law (ICAIL 2025), June 16–20, 2025, Chicago, IL, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3769126.3769208

1 Introduction

In 2016, the Malopolska Regional Parliament adopted the so-called anti-smog resolution for the city of Kraków - the former capital city of Poland. The legislative process arose due to the catastrophic state of air pollution in the city. The resolution effectively prohibited the use of coal and wood in heating installations, allowing only the use of gas and oil fuels. Polish law enables interested stakeholders to complain about regional regulations to the administrative court – and complaints were filed. The complainants argued that the regulation was not compliant with the law, in particular with the constitutional principles of private property protection and the principle of equal protection of rights. In fact, the adoption of the resolution effectively precluded the use of heating installations designed to use coal for heating purposes – a type of installation very popular in Poland during that time. Consequently, the complainants argued that the owners of coal-based installations would be discriminated against by the public authorities. Finally, the complainants raised the violation of the constitutional principle of proportionality.

In response, the Regional Parliament claimed that the resolution aimed at the protection of the constitutional principles of protection of human health and natural environment. The principle of proportionality had not been violated because the adopted prohibition was necessary to eliminate the smog problem; most importantly, however, the danger to human health and the natural environment caused by the intensity of the smog problem justified the adoption of total prohibition of coal and wood heating, rather than the adoption of more lenient regulations such as the use of special filters or fuel of higher quality. Eventually, the position of the Regional Parliament was upheld by the courts after a cassation complaint. This case is representative of many legal decisions (and other decisions) in which multiple aspects of a problem have to be balanced to reach a solution.

In the example, the factual situation was the passing of a regulation which introduced a prohibition of coal- and wood-based heating. This regulation was subject to evaluation regarding the applicable principles. Other situations can also be evaluated with respect to applicable principles. There may be situations composed of brute facts, for example, where a driver (or an autonomous vehicle) must decide how fast to drive to properly balance travel efficiency and safety. There may be a situation concerning interpretation and application of law, where a court may investigate a proper balance between legal certainty and predictability on the one hand and substantive fairness on the other. In this paper, we focus on situation type of the initial example, which arises in constitutional

*Michał Araszkiewicz contribution was funded by National Science Centre, Poland (grant no. 2023/49/B/HSS/007S3)

ACM ISBN 979-8-4007-1939-4/25/06
https://doi.org/10.1145/3769126.3769208

review in the US or the activities of the European constitutional courts such as the German Bundesverfassungsgericht. In fact, these legal contexts were crucial for the theoretical elaboration of the notion of legal principles: legal norms have to be balanced with each other in a specific context in order to produce an answer to a legal question [1]. Although the notion of legal principles and their application is widely discussed in jurisprudence, legal practice, and AI and Law literature (see Section 2), computational model of reasoning with legal principles warrants further development and elaboration, which is the aim of this paper.

The model supports the decision-making process in subject areas regulated by principles and, more generally, any areas where balancing of principles is involved. The history of principle-based reasoning in constitutional law, human rights protection, or EU law showcase the use of principles: they have to be balanced and the decision-maker should be looking for a solution that optimizes the realization of all relevant principles [1]. It is widely understood that principles in the sense discussed here are legal norms safeguarding legally relevant values [16]; and, importantly, each legally relevant value should be realized at least to a threshold which represents its *core*, that is, the minimal level of value realization that should be protected in any case, if possible. The general structure of arguments based on principles is known (for instance, [2, 5]). However, because of the high degree of generality and vagueness of the principles, it is often subject to debate whether one argument based on a principle is stronger than another; the application of principles involves a margin of discretion, which needs to be rationalised and justified [1]. In other words, to justify a decision based on principles, a qualitative, comparative judgment has to be made: one option weighs more (in terms of principles' satisfaction) than another. However, in order to more precisely support the decision-making process, our model provides a quantitative layer based on multi-criterial optimization techniques. This model serves an explanatory function and also enables experiments.

To focus the discussion, our worked example roughly corresponds to the issue of Kraków's anti-smog regulation with the addition of other local regulations. The issue to decide is to what extent these regulations are compatible with or comply with the legal principles on natural environment, equal protection, human health, and private property. The regulations must be considered in balance with the principles. Moreover, we claim that values underlie the principles, so that reasoning about principles and regulations amounts to reasoning about what values to champion.

Several formal models have been proposed to address issues such as those of the Kraków example (see Section 5). These approaches include (logical) reasoning with meta-rules, where conflicting rules are found and conflicts solved [17]. While rules in most cases will guide the decision effectively, where rules conflict, one has to look for the underlying principles. In our approach, we represent these principles as first-class citizens and explicitly represent the values that are promoted (demoted) if one adheres to these principles. We assume that these principles could be made explicit. We adopt the view, advocated in maintstream legal philosophy, that for each legally relevant value there exists a principle which supports or protects it [13]. Hence, instead of reasoning over rules, we reason over a multidimensional space of values associated with principles. Such reasoning is also common outside the field of law, for example, when deciding on business strategies, etc. In this paper, we propose and explore this approach.

The novel contribution of this paper is a formal framework to represent and reason with principles in proportionally optimised balance. The framework can be used: to make a decision such as in the case of the anti-smog law; or to evaluate regulations for compliance with principles, leading to the regulation being passed (dismissed) or upheld (overturned). Importantly, proportionality here implies some *fair* distribution of values according to the adjudicating authority; that is, one principle ought not unduly outweigh another principle without just cause.

In Section 2, we outline the elements of legal theory and practice with respect to rules and principles, scoping the issues. The formal model and worked example are presented in Section 3. The framework is discussed in Section 4, followed by related work in Section 5 and a conclusion in Section 6.

2 Rules and Principles in Legal Theory and Practice

In this section, we summarise the distinction between rules and principles in order to then focus on how principles are assessed relative to a situation and application [1, 2, 5, 9]. In particular, for principles, adjudicators must reason in a manner that uses balancing, whereas with rules, no such balancing is used.

To distinguish between rules and principles, the contemporary account considers the steps to be taken to determine which rule or principle is applicable in a given fact situation, thence what conclusions (in rules) or determinations (in principles) should hold.

Suppose a regulation that contains rules such as *Private cars are not allowed in the park* and *Safety-critical vehicles are allowed in the park*. For a rule-based regulation, we retrieve all rules wherein the premises of the rules are satisfied in the factual situation (are applicable to it) (*retrieval step*). Suppose the factual situation where the car is a private car being used for a party. If we see it in a park, then our example rule would appear to be applicable (*application step*). Generally for the applicable rules, the consequence holds - the vehicle violated the rule. However, rules may be *defeasible*; where there are *exceptions* and *defeaters*, the conclusion does not follow [21]. For example, suppose the private car was used in a safety-critical situation - the vehicle might have not violated the rule. Among the applicable rules, there may be some choice in the selection (*selection step*). For instance, if the consequences of applicable rules are incompatible, then it is necessary to apply meta-level preference relation between the rules, e.g., safety-criticality trumps recreational use. Such preference relations may be abstractly established, irrespective of the specificities of the fact situation. Furthermore, the decision-maker may attempt to restrict or extend the scope of a rule's application by means of interpretive moves [22, 26]. In legal decision-making, there may be some *meta* considerations about which rules apply, how they apply, and how they interact, but these are relatively restrictive. In any case, once the most preferred applicable rule has been identified, its conclusion is generally accepted as legally binding (*conclusion step*). Such rules are closest to logical rules and may be taken as representing the so-called *black letter* law, where the identification and application of legal rules are relatively straightforward, even if sometimes disputable.

For the application of principles, the retrieval step is similar to the application of rules. Previously, we introduced a case that raised several constitutional principles. In the above-mentioned fact situation, all principles were relevant for consideration. However, where the adjudicator is faced with claims of alternative and perhaps contrasting principles that may be applicable, the adjudicator must select which principles to apply in order to make a determination. In our work, we assume that principles have a conditional part describing the applicable circumstances.

In deciding which principles specifically to apply from the potentially applicable principles, the adjudicators identify the rationale for the principle, weigh the pros and cons of each, compare them to one another, and reach some decision that proportionately balances the values protected by the principles (*rationalisation step*). In this sense Alexy refers to principles as optimization commands: norms that requires something be realized to the greatest extent possible taking into account factual and legal limitations [1]. The degree of promotion or demotion of a principle is possible in specific situations. The complexity of the rationalisation step for principles contrasts with the crispness of predetermined preferences in the selection step for rules.

A key idea in our framework is that this rationale is provided by the *social values* that the principle aims to promote or demote. Promotion (demotion) relates to the circumstances which an agent advocates to bring about such that the value is recognised to hold (or not). These social values span the dimensions of the decision-making space. Given several principles, the issue is how to prioritise them in specific situations. In this sense, values are explicitly raised in relation to principles and reasoned with in the rationalisation step, while for rules, these may be implicit or not reasoned with as in the rationalisation step. The values associated to the principles can be viewed as fine-grained preference relations. This allows for reasoning about preferences in a multi-dimensional way which is different from the singular preference relation as applied in the rule-based approach. As the adjudicator considers values in relation to one another and relative to their importance, they are *balanced* with one another; principles, unlike rules, must be proportionally balanced to one another.

As in the example, the intensity of the smog caused such a danger to human health and the environment that it justified the adoption of a total prohibition of coal and wood heating; that is, the values underpinning the principles of human health and natural environment won over private property and equal protection of rights in this specific situation. Yet, the values were considered in proportional balance, since, after all, it was not decided that only human health matters and that there is no protection of private property. Rather, each applies to some greater or lesser extent and in relation to each other.

In an ideal situation, we could independently and maximally satisfy the value of each principle, but more often it is a matter of balancing between them - some values are more satisfied than others or not at all. To make a decision, we weigh the principles, finding an "optimal", balanced solution [1, 2, 4].

The main idea and key contribution of the paper is to provide a formal mechanism to represent an "optimal", multi-dimensional, proportionally balanced solution, applying principles and their associated values in a given factual situation to support a decision or help check compliance with principles. The key assumption (rooted in [27]) is that each principle can be represented along with a level of satisfaction of value(s). However, in [27], balancing principles was not sufficiently addressed.

In this paper we'll focus on principles, their associated values, and optimised balancing amongst them. We take the point of view of a singular adjudicator, who must decide whether a regulation complies with principles and to what extent. We leave for future work multiple agents having different values associated to principles, adhering to different principles or otherwise engaging is a dispute about principles.

3 Principles - Towards a Formalisation

In this section, we introduce the model of the reasoning with principles. The key element is a *principle compliance test*, the purpose of which is to test which of the available decisions should be made given applicable principles. This test has the following steps:

- Retrieval task, where the principles applicable to a particular situation are identified.
- Basic compliance task: analysis whether a given situation breaks the principle-based thresholds. If the goal of a principle is to promote a certain value, the basic test is to analyse whether a given situation promotes this value at all. This task can be done separately for every available decision and every applicable principle.
- Multi-objective optimisation task: all the available decisions and relevant principles are analysed and balanced to find the best one.
  This task requires taking into consideration not only all the applicable principles, but also comparison of available decisions in order to choose the best one (i.e., optimal with respect to applicable principles). To fulfill this task, a certain criterion of comparison is necessary (payoff function) as well as a mechanism to aggregate the levels of satisfaction of values by the decisions.

To realise this test, we detail the relevant components.

3.1 Basics

Some basic concepts:

- `Prop`, where each element is a proposition[1]. `Prop` = $\{f_x, f_y, ...\}$.
- `Value` = $\{v_x, v_y, ...\}$, where each element of `Value` is an abstract object that expresses a value concept such as *freedom*, *security*, etc.
- `Action` =$\{a_s, a_t, ...\}$, where each element of `Action` represents an action.

Our model of principles is based on the model of legal principles from [27]. Basically, principle can be understood as a minimal acceptable level of satisfaction of a certain value:

Definition 1 (Principle). *Let a pair* $p_x = \langle \mathcal{P}, \langle v_n, v_n^{min} \rangle \rangle$ *be a principle, in which* $\mathcal{P}$ *is a conjunction of propositions (*$\mathcal{P} \subseteq$`Prop`*),* x *is a proposition representing the name of principle,* v_n *is value name,*

[1]The proposal is neutral as to whether these are atomic or complex, or provide a semantic representation or are simply strings. The set of Prop can contain different sorts of propositions.

and v_n^{min} is a number representing a minimal level of satisfaction of value v_n. By P we denote a set of all principles.

$\mathcal{P}$, the conjunction of propositions, represents the conditional part of the principle, which represents the circumstances in which the principle is applicable, while $\langle v_n, v_n^{min}\rangle$ represents the goal of principle, that is minimum extent v_n^{min} to which the value v_n is promoted. The goal is not (relative to a conditional) a conclusion in a logical sense, i.e. it is not a truth bearer, but rather it is understood as guidance to an agent to make decisions that should promote the value to a certain level.

$\mathcal{P}$ can contain propositions which are, for the sake of example, a local regulation, a description of a situation, or some conjunctions of these. For simplicity, the principle assumes the minimal level of satisfaction of only one value.

There is a question concerning the source of v_n^{min}, that is, where to take this value from? Statutory principles do not introduce any number, assuming that lawyers can intuitively, on the basis of legal knowledge and experience, reconstruct such a minimum level of satisfaction of a value and use it in reasoning. In the introduction, we noted that lawyers do not use numbers in their judgments; rather, the levels can be taken as exact and calculable expressions of more intuitive and flexible notions of comparison. Moreover, the numbers can, to some extent, be (retrospectively) proposed on the basis of past judgments. Alternatively, they can be declared in advance by expert lawyers.

We have two functions to obtain the name of the value and minimal acceptable level of satisfaction of the value from the principle:

- $V : P \rightarrow [0..1]$.
  $V(p_\alpha) = v_n^{min}$ if $p_\alpha \in P \wedge p_\alpha = \langle \mathcal{P}, \langle v_n, v_n^{min}\rangle\rangle$.
- $VV : P \rightarrow Value$.
  $VV(p_\alpha) = v_n$ if $p_\alpha \in P \wedge p_\alpha = \langle \mathcal{P}, \langle v_n, v_n^{min}\rangle\rangle$.

The above definition may be illustrated by an example:

Example 3.1. In issuing a local regulation, four principles ought to be considered:

- $p_{privateP}$: it should protect private property
- $p_{equality}$: it should protect equal rights
- p_{health}: it should protect human health
- $p_{environment}$: it should protect the natural environment

These four principles aim at supporting four values: private property ($v_{privateP}$), equal rights ($v_{equality}$), health (v_{health}), and natural environment ($v_{environment}$). Below, we differentiate principles and values.

Next, we describe the relevant facts of the case, i.e., the institutional facts that determine the derivative outcomes of institutional reasoning, which are given as propositions of the applicable regulations such as:[2]

Example 3.2. Suppose a series of propositions:

- There is local regulation issued ($f_{localReg}$).
- There is no local regulation issued ($f_{noLocalReg}$).
- The local environmental regulation bans wood burning stoves ($f_{banWoodReg}$).
- The local environmental regulation obliges to filter wood burning stoves ($f_{filterWoodReg}$)
- The local environmental regulation does not contain any limitation of the usage of wood burning stoves ($f_{permWoodReg}$).

These propositions we take as atomic, e.g., these regulations exist and are applicable and guide the behavior of the addressees including (local) authorities. In this paper we do not consider how these regulations came into existence, nor their rationale, but we focus on how these regulations (and their associated principles) are used in reasoning about cases and whether the regulation in those case contexts abide these principles. We use $f_{localReg}$ to refer to a regulation, i.e. a set of rules that we consider with respect to the principle. All rules that $f_{localReg}$ consist of must satisfy the principle. In effect, the principle applies distributively over the rules in a regulation.

By $v_{privateP}^{min} = 0.6$ we denote a certain necessary threshold of private property protection; by $v_{equality}^{min} = 0.7$ a threshold of equal rights protection; by $v_{health}^{min} = 0.6$ a threshold of human health; and by $v_{environment}^{min} = 0.5$ a threshold for natural environment protection. Then the above principles may be formalised as:

- $p_{private} = \langle\{f_{localReg}\}, \langle v_{privateP}, 0.6\rangle\rangle$
- $p_{equality} = \langle\{f_{localReg}\}, \langle v_{equality}, 0.7\rangle\rangle$
- $p_{health} = \langle\{f_{localReg}\}, \langle v_{health}, 0.6\rangle\rangle$
- $p_{environment} = \langle\{f_{localReg}\}, \langle v_{environment}, 0.5\rangle\rangle$

The regulation must be considered with respect to the values of all of these principles, but this matters only where the principle is applicable, i.e. if the regulation contains any rule associated with that principle. That is, a health regulation may not pass an environmental principle because the regulation doesn't contain any rule associated with the environmental principle and hence the principle is not applicable to the regulation.

As we evaluate whether or not a regulation in a situation complies with a principle, we introduce regulatory situations:

Definition 2 (Situation). *Let $C = \{c_1, c_2, ...\}$ be a set of situations, where each situation c_i represents a set of propositions that determine/describe a situation ($c_\alpha \subseteq \mathtt{Prop}$). We assume that each proposition in c_i represents a fact; that is, every $f \in c$ represents a particular fact that holds in that situation.*

If $f_{banWoodReg} \in C_{banS}$ this should be interpreted as the rule $f_{banWoodReg}$ is applicable in the context C_{banS}.

Example 3.3 (cont.). For the environmental legislation case, the legislator could deliberate about three possible regulations: to issue a regulation without any limitations on the usage of wood burning stoves ($c_{noLImits}$), to issue a regulation with the prohibition of the usage of wood or coal burning stoves; or to issue a local regulation with a permission to use coal stoves with some constraints, i.e., a filter on the stoves. That is, $C = \{c_{banS}, c_{permS}, c_{noLimits}\}$, where:

- $c_{noLimits} = \{f_{permnWoodReg}\}$
- $c_{banS} = \{f_{localReg}, f_{banWoodReg}\}$
- $c_{permS} = \{f_{localReg}, f_{filterWoodReg}\}$

Regulations are chosen and evaluated relative to a context, considering the extent to which the regulations satisfy the values of the principles, which are associated with the regulations. With $f_{localReg}$ we allow that there are other regulations under consideration.

[2]This is a simplification of a case, which can be represented in a more sophisticated way with the use of complex propositions, weak/strong negations, etc.

Relative to a state of affairs, different principles may be applicable, so we need to identify those which apply.

Definition 3 (Applicable principles). *Let* $P_{c_x} = \{p \mid p = \langle \mathcal{P}, \langle v_n, v_n^{min} \rangle \rangle \wedge p \in P \wedge \mathcal{P} \subseteq c_x \wedge c_x \in C\}$ *be a set of principles applicable in the situation* c_x

A given principle is applicable in a certain state of affairs if this state of affairs (i.e. the propositions describing it) fulfills the conditional part of principle, by which we mean that propositions describing a factual situation include the propositions described in the conditional part of principle. More sophisticated forms of representation of fulfilling a principle's conditions, like using different interpretations, complex propositions, abstraction etc., are possible, but we leave this discussion for a future work.

Example 3.4. We have four principles: $p_{privateP}, p_{equality}, p_{environment}$, and p_{health}, each of them has the same conditional part ($\{f_{localReg}\}$) and two factual situations: c_{banS} and c_{permS}. Since $\{f_{localReg}\} \subseteq c_{banS}$ and $\{f_{localReg}\} \subseteq c_{permS}$, then all principles are applicable in both situations.

Different situations can promote different values to certain extents. In order to represent this, we evaluate to which level a given factual situation satisfies a given value:[3]

Definition 4. *Let* $VS : V \times C \longrightarrow [0..1]$ *be a function that, for a particular state of affairs and value, assigns a number representing the level of satisfaction of this value by this situation.*

Note that functions VS and V, although technically similar, represent different things: V returns the required acceptable level of satisfaction of value (on the basis of principle), while VS returns the extent to which a particular situation satisfies a given value. In other words: V represents what is preferred (in special cases what should be), while VS what is.

Example 3.5 (cont.). Consider situations: $c_{noLimits}, c_{banS}$ and c_{permS}, where:

- $VS(v_{privateP}, c_{noLimits}) = 0.9$
- $VS(v_{equality}, c_{noLimits}) = 0.8$
- $VS(v_{health}, c_{noLimits}) = 0.2$
- $VS(v_{environment}, c_{noLimits}) = 0.1$
- $VS(v_{privateP}, c_{banS}) = 0.7$
- $VS(v_{equality}, c_{banS}) = 0.8$
- $VS(v_{health}, c_{banS}) = 0.9$
- $VS(v_{environment}, c_{banS}) = 0.8$
- $VS(v_{privateP}, c_{permS}) = 0.8$
- $VS(v_{equality}, c_{permS}) = 0.9$
- $VS(v_{health}, c_{permS}) = 0.7$
- $VS(v_{environment}, c_{permS}) = 0.5$

Banning the usage of stoves satisfies the value private property to 0.7, equality to 0.8, human health to 0.9, and environment to 0.8. On the other hand, requiring filters (i.e. permission of the usage of stoves), the values are satisfied by 0.8, 0.9, 0.7, and 0.5, respectively. Note that, when there is a ban on the usage of stoves, the values v_{health} and $v_{environment}$ are promoted to a higher extent than when filters are permitted. On the other hand, when there is a ban on the usage of stoves, the values $v_{privateP}$ and $v_{equality}$ are lower than when filters are permitted. For a situation in which there are no limitations on the usage of stoves, $v_{privateP}$ and $v_{equality}$ are promoted to a very high level, but v_{health} and $v_{environment}$ are satisfied to a very low extent. These show how the values are promoted or demoted relative to the regulatory environments.

[3]This abstracts over interpretive variation, which we discuss in future work.

3.2 Threshold

Similar to [27], we understand a principle to establish a core threshold, that is, a minimal acceptable level of satisfaction of value. On the basis of this we can assume that a given situation is compliant with an applicable principle if the value level of the situation is equal to or greater than the threshold level required by the principle:

Definition 5 (Compliance with Principles). *A given factual situation* $c_x \in C$ *is compliant with relevant principles, which we denote by* $Comp(c_x)$ *if* $\forall p_\alpha \in P_{c_x}(VV(p_\alpha) = v_\beta \wedge V(p_\alpha) \geq VS(v_\beta, c_x))$

In other words, in order to be compliant with principles, a given situation must pass the filters set by all the relevant principles.

Example 3.6 (cont.). Now we consider whether the two described above situations (c_{banS} and c_{permS}) are compliant with the four available principles. In order to do this, we examine whether values promoted by the situations are above the thresholds set by the principles:

For $c_{noLimits}$:

- $VV_{p_{privateP}} = v_{privateP}, V_{p_{privateP}} = 0.6$, $VS(v_{privateP}, c_{noLimits}) = 0.9$ and $0.6 \leq 0.9$
- $VV_{p_{equality}} = v_{equality}, V_{p_{equality}} = 0.7$, $VS(v_{equality}, c_{noLimits}) = 0.8$ and $0.7 \leq 0.8$
- $VV_{p_{health}} = v_{health}, V_{p_{health}} = 0.6$, $VS(v_{health}, c_{noLimits}) = 0.2$ and $0.6 \nleq 0.2$
- $VV_{p_{environment}} = v_{environment}, V_{p_{environment}} = 0.5$, $VS(v_{environment}, c_{noLimits}) = 0.2$ and $0.5 \nleq 0.2$

For c_{banS}:

- $VV_{p_{privateP}} = v_{privateP}, V_{p_{privateP}} = 0.6$, $VS(v_{privateP}, c_{banS}) = 0.7$ and $0.6 \leq 0.7$
- $VV_{p_{equality}} = v_{equality}, V_{p_{equality}} = 0.7$, $VS(v_{equality}, c_{banS}) = 0.8$ and $0.7 \leq 0.8$
- $VV_{p_{health}} = v_{health}, V_{p_{health}} = 0.6$, $VS(v_{health}, c_{banS}) = 0.9$ and $0.6 \leq 0.9$
- $VV_{p_{environment}} = v_{environment}, V_{p_{environment}} = 0.5$, $VS(v_{environment}, c_{banS}) = 0.8$ and $0.5 \leq 0.8$

For c_{permS}:

- $VV_{p_{privateP}} = v_{privateP}, V_{p_{privateP}} = 0.6$, $VS(v_{privateP}, c_{permS}) = 0.8$ and $0.6 \leq 0.8$
- $VV_{p_{equality}} = v_{equality}, V_{p_{equality}} = 0.7$, $VS(v_{equality}, c_{permS}) = 0.9$ and $0.7 \leq 0.9$
- $VV_{p_{health}} = v_{health}, V_{p_{health}} = 0.6$, $VS(v_{health}, c_{permS}) = 0.7$ and $0.6 \leq 0.7$
- $VV_{p_{environment}} = v_{environment}, V_{p_{environment}} = 0.5$, $VS(v_{environment}, c_{banS}) = 0.5$ and $0.5 \leq 0.5$

In $c_{noLimits}$, although the environmental regulation, which does not limit the usage of stoves, promotes the values private property and equality to high levels, the values health and environment are below the threshold. Therefore this situation does not fulfill the minimal requirements of two principles and is not compliant with the principles as a whole. The other situations, c_{banS} and c_{permS}, all the relevant principles are satisfied (at least to the minimal acceptable level), so both situations are compliant with principles.

3.3 Optimisation

As pointed out in [3], principles are optimisation directives, which impose the choice of the best situation. Typically optimisation means the choice of the best possible option/decision/situation which maximises the payoff function. What, then, should be optimised in the principles' compliance testing?

Answering this question requires discussion of two important points:

- Firstly, the optimisation process relates to making a choice of the best possible situation (option/decision). On the basis of that we need a decision making mechanism, in which there will be a choice of a set of certain available decisions which bring about some situations amongst which the choice can be made.
- Secondly, a certain payoff function, which allows for comparison of different situations is necessary.

Below we present a discussion of both points. We start with a discussion of the payoff function, and then we introduce a decision-making mechanism.

It is important to emphasise (cf. [3]) that optimisation should be performed in the light of all the relevant principles. Therefore, we can assume that this should be a multidimensional optimisation, when each relevant value is represented by one dimension.

What to optimise? (the payoff function). Law does not explicitly define any payoff function, but we can try to reconstruct one. Every situation (potential decision) should be evaluated in the light of applicable principles. Each principle aims at promoting a certain value. These values are distinct and possibly conflicting (promoting one might demote another), hence the model should include a balancing mechanism. The goal is to introduce a mechanism of calculating the cumulated evaluation of the situation. Such an evaluation will be a basis for a comparison of the available decisions.

A number of choices should be made during this stage. Firstly, it should be decided what should be cumulated: (1) the absolute level of value promotion for each dimension in the value space, or (2) alternatively, the surplus over the thresholds?
Since the actual bandwith of appropriate actions is the interval between the minimum value of a particular dimension v_n^{min} and 1 (as all dimensions are normalized), the surplus gives a more intuitive indication of the 'effects' of a regulation (assuming that the intended effects are actually achieved by the people adhering to that regulation). We could compare the effects of different regulations by looking at the respective surplus effectuated. Another aspect in balancing principles is balancing between value dimensions. The Law emphasizes the role of proportionate promotion of values in principles, however some values may be more or less important than others. In Law the necessity of proportionality and 'fair distribution' of values is emphasized. If we consider absolute levels of values' promotion as a ground of optimisation, it will be not possible to evaluate whether the distribution is 'fair' amongst all values, while if we take into consideration surpluses, then we can optimise not only the values' promotion, but also the 'fair' distribution of surpluses over the thresholds.

On the basis of the above some auxiliary concepts should be introduced:

Definition 6 (Difference). *Let $Diff : P \times C \rightarrow [0..1]$ be a difference function, $Diff(p_\alpha, c_x)$, which returns how much the level of satisfaction of value v_β by situation c_x is above the threshold for that value given in the principle p_α and applicable in factual situation c_x:*
If $p_\alpha \in P_{c_x}$, $VV(p_\alpha) = v_\beta$, and $Comp(c_x)$ then $Diff(p_\alpha, c_x) = VS(v_\beta, c_x) - V(p_\alpha)$

This can be illustrated by an example:

Example 3.7 (cont.). Differences between the levels of satisfaction of particular values by a given situation and the thresholds set by applicable principles:
For c_{banS}:

- $Diff(p_{privateP}, c_{banS}) = VS(v_{privateP}, c_{banS}) - V(p_{privateP}) = (0.7 - 0.6) = 0.1$,
- $Diff(p_{equality}, c_{banS}) = 0.8 - 0.7 = 0.1$
- $Diff(p_{health}, c_{banS}) = 0.9 - 0.6 = 0.3$
- $Diff(p_{environment}, c_{banS}) = 0.8 - 0.5 = 0.3$

For c_{permS}:

- $Diff(p_{privateP}, c_{permS}) = (v_{privateP}, V_{p_{privateP}}) = (0.8 - 0.6) = 0.2$,
- $Diff(p_{equality}, c_{permS}) = 0.9 - 0.7 = 0.2$
- $Diff(p_{health}, c_{permS}) = 0.7 - 0.6 = 0.1$
- $Diff(p_{environment}, c_{permS}) = 0.5 - 0.5 = 0$

As we already mentioned, there is a kind of meta-principle requiring the proportional balancing of principles [1]. In other words, the optimisation should not only concern maximisation of a value's promotion, but also should take into consideration "fair distribution" of surplus over the thresholds, to avoid a situation in which one value is promoted to a very high extent, while the others are just above their thresholds. Moreover, such an understanding of a proportional optimisation is also important in the light of human perception: increasing the level of satisfaction of a value which is already promoted to a very high level (e.g. from 0.9 to 1.0) does not have a strong influence on human satisfaction, while a similar change for a value which is only just above the threshold can significantly increase the personal satisfaction. This can be illustrated by an example of buying new home: increasing surface from 50 to 70 square meters significantly changes the life comfort, while changing from 220 to 240, not necessary. In order to include the proportionality aspect into the model, a non-linear function using values $Diff$ as argument should be added to the model. The non-linear function of the formula should fulfill some properties:

- It should be continuous.
- It has to be monotonically increasing (higher $Diff$ should cause higher value of the function's output).
- Lower values of $Diff$ should have a stronger influence on the function's output than higher ones.

There are a number of functions which can correctly represent the above properties. For our model, we use the logarithmic function:

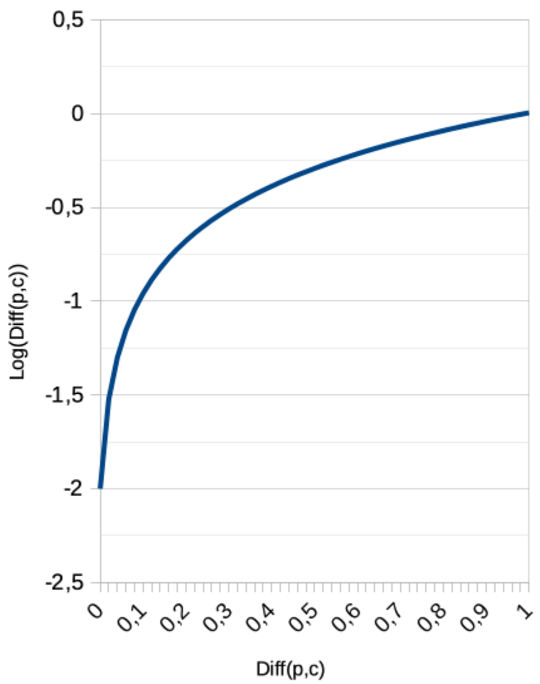

Figure 1: Shape of function $log_b(Diff(p_\alpha, c_x) + 0.01)$

$log_b(Diff(p_\alpha, c_x) + 0.01)$ (see fig: 1). Constant (0.01) is necessary to allow calculating a surplus if $Diff(p_\alpha, c_x) = 0$.

On the basis of the above function, we can analyse the surplus over the threshold declared in principle:

Definition 7 (Cumulated surplus). *Let $Sur : C \longrightarrow \mathbb{R}$ be a function returning the number representing the cumulated surplus over the thresholds declared in principles applicable for the situation c_x. If $comp(c_x)$ then $Sur(c_x) = \frac{\sum_{p_\alpha \in P_{c_x}} (log_b(Diff(p_\alpha, c_x)+0.01))}{n}$*

The base of the logarithm (b) in the above function should be greater than 1. The higher the base, the more non-linear the formula will be. In our model $b = 10$, but other bases are also possible. n is the number of applicable principles. Dividing the sum by the number of principles keeps the evaluation independent of the number of applicable principles. Although, the result of $log_n(Diff(p_\alpha, c_x) + 0.01)$ is mostly negative, it is monotonically increasing (higher $Diff(p_\alpha, c_x)$ causes, ceteris paribus, higher $Sur(c_x)$). Note that since we assumed that $Diff$ is calculated for situations which are compliant with principles (are above the threshold), then $Diff$ is always positive and between 0 and 1.

Example 3.8 (cont.). On the basis of already calculated values of $Diff$ function, we can calculate a cumulated surplus:

- $Sur(c_{banS}) = \frac{log_{10}(0.11)+log_{10}(0.11)+log_{10}(0.31)+log_{10}(0.31)}{3} = -0.978163747338335$
- $Sur(c_{permS}) = \frac{log_{10}(0.21)+log_{10}(0.21)+log_{10}(0.11)+log_{10}(0.01)}{3} = -1.43805624179131$

Decisions. The compliance with principles' testing is usually performed when there is a necessity to make a certain choice, perform an action, or make a certain decision which can potentially be or not compliant with principles. Moreover, whatever decision on what action to take ought (ideally) to be one which is proportionally optimal with regards to the principles and their values.

In order to represent this we introduce a very simple change state mechanism in which performing an action changes the situation:

Definition 8 (State changes). *Let $T : C \times A \longrightarrow C$ be a partial function which represents changing of states: by $T(c_\alpha, a_\beta) = c_\gamma$ we represent that performing action a_β in a context (situation) c_α changes situation to c_γ*

Example 3.9 (cont.). In order to illustrate the concept of state changes, we have to introduce additional situation. Let $c_{noLocalReg} = \{f_{noLocalReg}\}$ be a situation in which there is no local regulation concerning environment in the city. We also assume that $VS(v_{environment}, c_{noLocalReg}) = 0.2$, which means that the level of satisfaction of value $v_{environment}$ by situation $c_{noLocalReg}$ is below the threshold (0.5), thanks to which it is not compliant with principle $p_{environment}$.

On the basis of the above we assume that there are two possible actions that can be taken: to issue a regulation with a total ban on the usage of stoves, or to issue an environmental regulation without such a strong ban, but with an obligation of using filters. We can represent this as:

- $\texttt{Action} = \{a_{issue_{banS}}, a_{issue_{oblFilters}}\}$

We can assume that:

- $T(c_{noLocalReg}, a_{issue_{banS}}) = c_{banS}$
- $T(c_{noLocalReg}, a_{issue_{oblFilters}}) = c_{permS}$

In other words, issuing a regulation with a ban brings about the situation in which the usage of stoves will be forbidden, while issuing a regulation with the obligation of using filters brings about the situation with a regulation allowing for the usage of stoves.

Function T returns the results of the action made by an agent. The agent's behavior and the results of action can differ depending on the context (state the agent is in). In order to represent this we introduce the concept of decision which is an action made in a certain context:

Definition 9. *Let $D = \{d_1, d_2, ...\}$ be a set of decisions, which are tuples $< c_\alpha, a_\beta >$, where c_α is a set of propositions that describe a context and a_β is the action executed relative to the context, where every $c_\alpha \subseteq Prop$ and every $a_\beta \in Action$.*

Since T is a partial function, not every situation is available from every situation. On the basis of the above for every situation we can define a set of best possible available decisions.

Example 3.10 (cont.). In a situation $c_{noLocalReg}$ there are two possible decisions $D = \{d_{issueBanS}, d_{issuePermS}\}$, where:

- $d_{issueBanS} = \langle c_{noLocalReg}, a_{issue_{banS}} \rangle$
- $d_{issuePermS} = \langle c_{noLocalReg}, a_{issue_{oblFilters}} \rangle$

Out of the available possible decisions, we want to select that which is the "best decision", which ought to be that which corresponds to the outcome that is proportionally optimal.

Definition 10 (Best decisions). *By $\texttt{BestDecisions}_{c_\gamma}$ we denote a set of decisions which are available in situation c_γ and compliant with the binding principles.*

A given decision will belong to $\texttt{BestDecisions}_{c_\gamma}$ *if it is compliant with principles, and there is no other decision which brings about the situation with a higher cumulated surplus:*

$\texttt{BestDecisions}_{c_\gamma} = \{d_\alpha \mid d_\alpha = \langle c_\gamma, a_\mu \rangle \wedge T(c_\gamma, a_\mu) = c_\beta \wedge \neg\exists c_\beta s.t. Sur(c_\beta) > Sur(c_\gamma))\}$

In other words, the best decisions are decisions which are compliant with principles, and they bring about the maximal cumulated surplus.

Note that there could be more than one best decision: it is not impossible that two or more available decisions will have the same cumulated surplus. Moreover, since the aim of this paper is to introduce the mechanism of testing the compliance with principles, we do not introduce any decision-making mechanism. The definition of the best decision which we have introduced above is not a decision making mechanism, but the best decision in the light of applicable principles. The decision maker, can have other reasons which may lead to make suboptimal (w.r.t. principles) choice.

Example 3.11 (cont.). In situation $c_{noLocalReg}$, there are two possible decisions $d_{issueBanS}$ and $d_{issuePermS}$ which bring about situations c_{banS} and c_{permS}. In previous examples we show that both are compliant with principles, but since principles require not only being compliant, but also being optimal (the choice of the best possible decision with respect to the applicable values), we have to compare them in the light of cumulated surplus:
Since $T(c_{noLocalReg}, a_{issue_{banS}}) = c_{ban_S}$,
$d_{banS} = \langle c_{noLocalReg}, a_{issue_{banS}} \rangle$, and $Sur(c_{banS}) > Sur(c_{permS})$
then $\texttt{BestDecisions}_{c_{noLocalReg}} = \{d_{banS}\}$
In other words, the best decision, in the light of applicable principles, is to issue a regulation with ban on the usage of stoves, because the cumulated surplus of forbidding the usage of stoves is higher than cumulated surplus of allowing the usage of stoves, but with filters. And so it was decided in the narrative of our working example in Section 1.

4 Discussion

The computational model presented in this paper captures essential legal intuitions and provides a framework for addressing complex decision-making scenarios involving conflicting principles. One of the foundational aspects of the model is its ability to determine the relevance of principles to specific fact situations. This determination depends on the applicability conditions and contextual elements described within the given situation. Another critical feature of the model is its representation of principles as satisfiable to varying degrees. This approach reflects the reality that legal principles are rarely absolute; instead, they can be promoted or demoted depending on the context. To achieve this, the model assigns values between 0 and 1 to measure the extent to which a principle is satisfied. Furthermore, the introduction of thresholds ensures that no solution is deemed acceptable if it fails to satisfy a principle beyond a minimum acceptable level. These thresholds embody a fundamental legal intuition — that certain degrees of a principles' realisation hold non-negotiable importance within the decision-making process.

The proportional balancing of principles in this model relies on the use of logarithmic functions, which serve two purposes. First, logarithmic scaling encapsulates the idea of diminishing returns, acknowledging that additional efforts to promote a principle yield progressively smaller benefits. This feature is particularly relevant in legal contexts where resources or competing values must be distributed efficiently. Second, logarithmic functions promote an equitable distribution of satisfaction among principles. By tempering the focus on maximizing the realization of any single principle, the model ensures that no principle is excessively prioritized at the expense of others. This balancing mechanism aligns with the broader aim of proportionality in legal reasoning. The reliance on numerical methods and mathematical functions in the model may seem unconventional to legal practitioners. Lawyers do not explicitly quantify principles or employ mathematical functions in their reasoning. However, these methods are incorporated not to simulate the communicated legal reasoning but to enhance its computational tractability and transparency.

Our model assigns specific numbers to the degree of promotion or infringement of principles and uses logarithmic functions to represent how these degrees change in different situations. Obviously, this mathematical formalism does not serve any direct role in justificatory reasoning, but it may be regarded as providing a quantitative explanation for a decision, representing sets of numbers compatible with this decision. The exact numbers used in our modeling could be different, but they are not entirely arbitrary, because they have to be found in specific ranges in order to remain compatible with an outcome. We do not claim that quantitative explanatory "theories" yielded by our model are "true" in any sense. Indeed, in general philosophy of science, it is well-known that any set of evidence may be consistent with an infinite number of theories explaining it, a phenomenon known as underdetermination of scientific theory [8]. However, we claim that our quantitative, explanatory approach may enhance the work of decision-making support models by making reasoning more precise and transparent and offering a possibility to test and refine the "theory" against different cases and examples.

Numerical representations allow the model to articulate the assumptions underlying its conclusions, enabling users to scrutinize and adjust those assumptions as necessary. The inclusion of adjustable parameters, such as thresholds and alternative functions, provides users with the flexibility to tailor the model to specific legal contexts or policy objectives. To make an analogy, creating a robot that coordinates eye-hand coordination to throw a ball to hit a moving object does not imply that all of the robot's maths and mechanics are found in the person throwing the ball. Rather, the robot is a model of some aspects of the human's capability, but it nonetheless helps progress understanding of what is involved in a human's capability in a way not otherwise available.

By integrating these features, the model offers a practical and transparent framework for supporting legal decision-making. It not only facilitates the assessment of principle compliance but also enables stakeholders to experiment with different configurations to explore alternative solutions. This adaptability ensures that the model can be employed across a wide range of scenarios, from judicial adjudication to policy analysis and automated legal reasoning systems. In doing so, it bridges the gap between the abstract nature of legal principles and the demands of practical application, laying the groundwork for a new generation of computational tools in Law.

5 Related Works

Although there are a number of papers discussing the issue of compliance with (legal) rules, [6, 7, 10, 15, 18], they do not introduce any principle-like concept. Other than the jurisprudential literature, our work is cognate to [20], where the author introduces the concept of goal-norms which can be understood as principles, and [27], where the formal concept of principles with a reasoning mechanism were introduced. Sartor [20] divides norms into:

- action-norms, requiring the full accomplishment of a certain action or omission (as a matter of obligation, or as the condition for the validity of a certain act);
- goal-norms, requiring the appropriate pursuit of certain objectives.

Sartor adopts the view that objectives are expressed by levels to which some values should be promoted [20]. Hence Sartor distinguishes two border values of promotion: *core threshold* and *maximum satisfaction threshold*.

A so-called "core" threshold is the level of satisfaction of a certain value that should never be violated, while a maximum satisfaction threshold is the level where further realization of a given value is not practically possible. Sartor also observes that the level of satisfaction with value promotion is often non-linear, as shown on the example of purchasing an apartment: ".. the additional benefit obtained by moving from 20 to 40 square meters is usually very important, the additional benefit obtained by moving from 200 to 220 is likely to be less significant.". In our work, the core threshold is declared in the principle, while maximum satisfaction as well as non-linear character of satisfaction is represented by a logarithm in a cumulated surplus formula. Our work extends the [20] work by introducing a concrete formal framework, which allows for testing the compliance with principles and a formal mechanism allowing for optimisation, which is (following [3]) crucial for the reasoning with principles.

Sartor's work [20] was the basis of [27], where the authors introduce the formal model of reasoning with principles. Our model and [27] share the concept of principle and the mechanism of basic compliance with principle testing (thresholds on values), but [27] does not take into consideration the optimisation and proportionality aspects of compliance testing.

A different concept of legal principles is presented in [24]. On the basis of Reason-Based Logic [11], [24] assumes that the conclusion of a principle is a term with free variables that is the reason for a given state of affairs. Our model allows us to particularize the concept of reason: by viewing reason as the minimum desired level of value promotion we are able to make decisions based on legal principles. Although [24] discuss the issue of balancing, they do not introduce any balancing mechanism, assuming the predicate *outweighs* to represent that certain reasons are stronger than others. The model does not introduce any mechanism to obtain such a knowledge. Furthermore, [24] observe that a significant difference between rules and principles can stem from the fact that rules are considered in isolation, whereas a principle reveals its character while interfering with other rules or principles. Our optimisation mechanism implements this property, integrating all the applicable principles into one comparison metrics.

A model of balancing in Law has been introduced in [25]. The authors, on the basis of the Art.7 and 8 of EU Charter of Fundamental Rights, discuss the balancing of right to privacy and access to information. In contrast to our work, this model is domain dependent: the authors introduced a model of one concrete balancing without an attempt to generalize their approach to other objects of balance. They do not introduce principles, but they focus on particular problem of relation between right to privacy and access to information. Their model is build on the analysis of proportion between the levels of realisation of these two concepts plus some auxiliary, domain specific components like status of the analyzed persons, sphere of information, or time. In contrast to this paper, their model does not discuss what balancing between a number of values nor non-linear aspects of proportionality; the *outcome* (variable u in this model) expressing a relation between levels of satisfaction of right to privacy and access to information has purely linear character, while this is not necessarily the case (see sec. 2).

Balancing is discussed in the [14], where the authors present a mechanism allowing for extracting rules on the basis of balancing values. The purpose of the model is, however, significantly different than ours. The model from [14] aims to introduce a mechanism which allows for extraction of rules on the basis of balancing of values, while our model aims at testing the compliance with legal principles, which is not discussed in [14]. Despite different purposes, their mechanism share some common elements with ours. In particular, both models have the possibility of balancing between more than two values. However, the details are different yet. In [14], the balancing formula is built on the assumption of a linear relation between the levels of promotion (or demotion) of values, while our model takes into consideration non-linear relations between the levels of satisfaction of values.

The concept of balance is central in Lauritsen's proposal [12] where the author introduces a quantitative "choiceboxing" approach towards balancing of reasons through calculation of their weights. However, the author does not invoke the optimization mechanism and the concept of core thresholds, which means that the referred model is less directly tailored to the representation of reasoning with principles as they are understood in legal theory.

Finally, our proposal has connections to the idea of quantitative approach to theory coherence in legal reasoning. In [23] Bench-Capon and Sartor mapped the qualitative elements of theories of legal cases (such as factors and values) to the numbers which then represent the parameters of a neural network, to enable experimentation. The model developed here focuses on a numerical representation and less on the qualitative reasoning, however, it clearly enables a mapping from the numerical representation to the qualitative, comparative argumentation concerning promotion of principles the lawyers naturally engage in. The numerical representation of reasoning with principles may be seen as a quantitative "theory" of a case explaining the result, which opens the possibilities to integrate our results with coherentist stream of AI and Law research and reinforce the debate on theory coherence metrices.

6 Conclusion and Future Work

This paper introduced a formal framework for reasoning with legal principles, focusing on proportional optimisation to balance values that collide in specific situations. By treating principles as thresholds of value satisfaction and employing quantitative methods, we demonstrated how multi-objective optimisation can aid decision-making in complex legal and regulatory contexts. Using the Kraków anti-smog case as a central example, the study showcased how competing values—such as human health and environmental protection on the one hand, and property rights and equality on the other—can be balanced to arrive at reasonable and justified outcomes.

While the use of numbers and mathematical functions does not directly replicate the reasoning or arguments typically employed in legal contexts, they serve an important explanatory purpose by illustrating exemplary degrees of value satisfaction compatible with the outcome. These quantitative tools enhance the precision of reasoning, making it more transparent and enabling the incorporation of computability into decision-making processes. This computational perspective does not replace the intuitive and logical-argumentative layers of legal reasoning but rather complements them.

The proposed model embodies and integrates numerous intuitions present in preceding approaches to reasoning with principles, unifying them into a computable framework. This synthesis offers a mechanism to support adjudicators and other decision-makers in crafting proportionate and well-justified decisions.

There is a number of directions in which our model can be developed. One of them is its application into the reinforcement learning-driven, autonomous devices which have to operate in a legally regulated environment. Our model can be used as a component of a reward function[4].

The model also can be further developed to investigate the interplay between qualitative argumentation and quantitative optimization in modeling reasoning with principles. As noted above, the exact numbers in our modeling are used as examples, but they are not entirely arbitrary because they have to belong to specific ranges in order to remain compatible with the result. We intend to apply our model to larger datasets of cases to investigate how different outcomes will constrain the ranges of numbers used to represent the degrees of principles' promotion and demotion. Such operations may be interpreted as revisions of initial theories of cases guided by the ideal of coherence. Specifically, we intend to apply the model to represented disputed issues in legal interpretation, where decisions are often backed by balancing of values but the balancing is not explicitly expressed in the rationales of judicial decisions.

References

[1] Robert Alexy. 2002. *A theory of constitutional rights.* Oxford University Press.
[2] Robert Alexy. 2003. On Balancing and Subsumption. A Structural Comparison. *Ratio Juris* 16, 4 (2003), 433–449.
[3] Robert Alexy. 2017. Proportionality and Rationality. In *Proportionality: New Frontiers, New Challenges*, Vicki C. Jackson and MarkEditors Tushnet (Eds.). Cambridge University Press, 11–12.
[4] Michal Araszkiewicz. 2010. Balancing of Legal Principles and Constraint Satisfaction. In *Legal Knowledge and Information Systems - JURIX 2010: The Twenty-Third Annual Conference on Legal Knowledge and Information Systems, Liverpool, UK, 16-17 December 2010.* 7–16.
[5] Michał Araszkiewicz. 2015. *Argument Structures in Legal Interpretation: Balancing and Thresholds.* Springer International Publishing, Cham, 129–150.
[6] Trevor Bench-Capon and Sanjay Modgil. 2017. Norms and value based reasoning: justifying compliance and violation. *Artificial Intelligence and Law* 25, 1 (2017), 29–64.
[7] B Burgemeestre, J Hulstijn, and Yao-Hua Tan. 2010. Value-Based Argumentation for Justifying Compliance. In *Deontic Logic in Computer Science*, Guido Governatori and Giovanni Sartor (Eds.). Lecture Notes in Computer Science, Vol. 6181. Springer Berlin Heidelberg, 214–228.
[8] Pierre Duhem. 1954. *The Aim and Structure of Physical Theory.* The Princeton Press.
[9] R. Dworkin. 1978. *Taking Rights Seriously. New Impression with a Reply to Critics.* Duckworth.
[10] Enrico Francesconi and Guido Governatori. 2022. Patterns for legal compliance checking in a decidable framework of linked open data. *Artif. Intell. Law* 31, 3 (July 2022), 445–464.
[11] J. Hage. 1995. Teleological reasoning in reason-based logic. In *Proceedings of the 5th international conference on Artificial intelligence and law (ICAIL '95).* ACM, New York, NY, USA, 11–20.
[12] Marc Lauritsen. 2015. On balance. *Artif. Intell. Law* 23 (2015), 23–42.
[13] N. MacCormick. 2005. *Rhetoric and the Rule of Law.* Oxford University Press.
[14] Juliano Maranhão, Edelcio G. de Souza, and Giovanni Sartor. 2021. A Dynamic Model for Balancing Values. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law* (São Paulo, Brazil) *(ICAIL '21).* Association for Computing Machinery, New York, NY, USA, 89–98.
[15] Sanjay Modgil, Nir Oren, Noura Faci, Felipe Meneguzzi, Miles Simon, and Michael Luck. 2015. Monitoring compliance with E-contracts and norms. *Artificial Intelligence and Law* 23, 2 (June 2015), 161–196. https://hal.science/hal-01212164
[16] A. Peczenik. 1996. Jumps and Logic in the Law. *Artificial Intelligence and Law* 4 (1996), 296–329.
[17] Henry Prakken and Giovanni Sartor. 2015. Law and logic: a review from an argumentation perspective. *Artificial Intelligence* 227 (2015), 214 – 245.
[18] Livio Robaldo, Sotiris Batsakis, Roberta Calegari, Francesco Calimeri, Megumi Fujita, Guido Governatori, Maria Concetta Morelli, Francesco Pacenza, Giuseppe Pisano, Ken Satoh, Ilias Tachmazidis, and Jessica Zangari. 2023. Compliance checking on first-order knowledge with conflicting and compensatory norms: a comparison among currently available technologies. *Artif. Intell. Law* 32, 2 (June 2023), 505–555.
[19] Manel Rodriguez-Soto, Roxana Radulescu, Juan A. Rodriguez-Aguilar, Maite Lopez-Sanchez, and Ann Nowe. 2023. Multi-objective reinforcement learning for guaranteeing alignment with multiple values. 2023 Adaptive and Learning Agents Workshop at AAMAS, ALA 2023.
[20] Giovanni Sartor. 2010. Doing justice to rights and values: Teleological reasoning and proportionality. *Artificial Intelligence and Law* 18 (6 2010), 175–215. Issue 2.
[21] Giovanni Sartor. 2012. Defeasibility in Legal Reasoning. In *The Logic of Legal Requirements: Essays on Defeasibility.* Oxford University Press.
[22] Antonin Scalia and Bryan A. Garner. 2012. *Reading law : the interpretation of legal texts / Antonin Scalia and Bryan A. Garner.* Thomson/West, St. Paul, MN.
[23] Giovanni Sartor Trevor Bench-Capon. 2001. A Quantitative Approach to Theory Coherence. In *Legal Knowledge and Information Systems - JURIX 2001: The Fourteenth Annual Conference.* 53–62.
[24] Bart Verheij, Jaap C. Hage, and H. Jaap Van Den Herik. 1998. An Integrated View on Rules and Principles. *Artificial Intelligence and Law* 6, 1 (01 Mar 1998), 3–26.
[25] Frederike Zufall, Rampei Kimura, and Linyu Peng. 2023. Towards a simple mathematical model for the legal concept of balancing of interests. *Artificial Intelligence and Law* 31 (12 2023), 807–827. Issue 4.
[26] Tomasz Zurek and Michał Araszkiewicz. 2013. Modeling Teleological Interpretation. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law* (Rome, Italy) *(ICAIL '13).* Association for Computing Machinery, New York, NY, USA, 160–168.
[27] Tomasz Zurek, Michał Araszkiewicz, and Dorota Stachura-Zurek. 2022. Reasoning with principles. *Expert Systems with Applications* 210 (2022), 118496.

[4][19] discusses a similar approach, but not related to the principles and with a different formal model