Interpretive Normative Systems

Juliano Maranhão [1]

University of São Paulo Law School and Center for AI- C4AI/USP
Largo São Francisco, 92- São Paulo- Brasil

Giovanni Sartor

CIRSFID- University of Bologna and European University Institute
Via Galliera, 3, Bologna, Italy

Abstract

We provide a formal definition of normative systems, which is compatible with different conceptions of the relation between law and morality. We embed a model for balancing values into an architecture of i/o logics representing conceptual, deontological and axiological rules. In particular, we provide a formal representation of three versions of the so-called Radbruch's formula, according to which legal obligations hold unless they reach a certain degree of immorality. Accordingly, we define eight different entailment relations, which correspond to eight different legal theories concerning the relation between law and morality.

Keywords: normative systems, input/output logics, balancing values.

1 Introduction

The regulation of human action has two sides. On the one side it aims to achieve certain values, i.e., goals that are socially desirable. Such values may consist in individual entitlements or rights (e.g. freedom of speech, property, privacy) or collective/social objectives (e.g. public health, national security, etc.). On the other side, the regulation specifies that certain actions may or may not be accomplished under certain antecedent conditions. The first is the dimension of consequentialism (also called teleology or axiology), according to which actions are evaluated according to their future impact on the relevant values: they are prohibited if they have a negative aggregated impact on the relevant values and they are permitted otherwise. The second is the dimension of deontology, according to which actions are evaluated according to the context in which they were accomplished: they are impermissible (or respectively permissible) if they are accomplished under conditions that trigger, through a rule, their prohibition (respectively permission).

[1] This research is developed within the Project FAPESP 2019/07665-4, Natural Language Processing Group of the C4AI-USP.

The two dimensions should ideally be aligned, since the deontological rules in the regulation are meant to serve the values aimed at by the regulation. The alignment is successfully achieved when the circumstances under which rules prohibit (permit) an action correspond to the circumstances under which the action would be detrimental (favourable) to the relevant values. However, a mismatch is also possible: what is deontologically prohibited may be axiologically required (having a positive impact on the relevant values) and what is deontologically permitted may be axiologically prohibited. For simplicity's sake, we assume that axiological components only pertain to political morality, while deontological components only pertain to positively enacted law. However, as we shall remark later, our approach can also deal with the incorporation of axiological components in the positively enacted law.

A long standing problem in legal theory concerns exactly the criteria for the identification of valid law in case of mismatch between law and morality. In the contemporary debate, the difficulty rests on how to sustain the authority of legal rules while excepting their application when it would lead to morally unacceptable results.

Non-positivist theories, such as those put forward by R. Dworkin [5] and R. Alexy [2], affirm a necessary but nuanced relation between law and moralty: on the one hand legal interpretation and argumentation may include evaluative efforts meant to align deontology and axiology; on the other hand, in some cases, immorality may entail legal invalidity. In particular, Alexy refers to the formula originally proposed by Gustav Radbruch [13] to determine the (in)validity of Nazi's laws: *laws enacted by proper authority and power are legally valid unless they reach an unberable degree of immorality or injustice.*

Positivist theories on the other hand, reject the view that necessarily the identification of law is dependant on moral considerations, while accepting that the immorality of a law may justify its modification or even the refusal to apply it, when this would lead to morally unacceptable consequences [8].

In our framework, the identification of the obligations and permissions derived from the normative system vary according to the version of the Radbruch's formula assumed, which, in its turn, reflects a particular conception about the morality of Law.

Our effort has not only a theoretical import for legal philosophy, but also a practical import for the design of intelligent normative agents. In a human-centered AI, artificial agents must not blindly apply predefined rules, but also be able to determine how best to apply such rules and even refrain from complying with them when that might offend the underlying social values and individual rights.

In Section 2 we introduce the normative sets of conceptual, modulation, deontological and axiological rules. In Section 3 we introduce a logical architecture of i/o logics based on our concept of Normative Systems. In Section 4 we define an operator of axiological entailment. In Section 5 we introduce the concept of normative theories and we identify eight different legal theories regarding the relations between law and morality, based on three versions of

Radbruch's formula.

2 Normative Sets

We shall use the term *"normative set"* to refer to sets of different kinds of rules: (i) a set of *conceptual rules*; (ii) a set of *modulation rules*; (iii) a set of *deontological rules* (iv) a set of *axiological rules.*

Conceptual rules consist in the ascription of a legal meaning or concept, *i.e.* they state that the entities described by certain factors *count as* (are to be classified as) instances of the ascribed concept (see [7]). We represent conceptual rules in the form (a, c) where a is the triggering factor (or conjunction of factors) and c is the ascribed concept. For instance, a conceptual rule stating that a message exchange stored in a mobile phone (sms) counts as "data" can be represented as (sms, dat).

Modulation rules specify the extent to which the presence of a factor affects the impact of actions on values. Such modulations reflect both causal connections (that the action, given the factor is likely to produce a certain individual or social outcome) and evaluative assessments (that the outcome of the action will count as an impact on the value). The values may consist in individual or social rights, moral principles, or collective goals.

We distinguish three kinds of modulation rules: baseline, intensifier and attenuator rules (following [4]). A baseline rule specifies that an action has a certain impact on a value, in the absence of relevant circumstances. That is, baseline modulation rules are those pairs where the body is a tautology, while intensifiers and attenuators are rules where the body is non-tautological. An intensifier rule specifies that the presence of a factor (the intensifier) increases the action's impact on the value (its index is positive). An attenuator rule specifies that the presence of a factor (the attenuator) decreases the action's impact on the value (its index is negative). We represent modulation rules in the form $(a, V^x)_i$, where a is the triggering factor, V is the affected value, x is the action at stake, and i is the extent of the modulation. For a baseline example, consider the rule specifying that the action consisting in the access to any item in a search by the police demotes the value of Privacy to the extent 0.2, which we model as $(\top, Privacy^{acc})_{.2}$. For an intensifier, consider the rule that the impact of this action on privacy is increased if the item is a mobile phone $(mob, Priv^{acc})_{.8}$. For an attenuator, consider that the impact is decreased if the mobile phone is not personal $(\neg pers, Priv^{acc})_{-.4}$.

We distinguish two kinds of rules establishing obligations or permssions, deontological and axiological ones. Such rules lead to deontic conclusions which may be in conflict.

Deontological rules link the (deontological) prohibition or permission of a given action to the presence of certain antecedent conditions. We model deontological rules in the form (a, x), where a is the triggering factor (or concept) and x is the obligatory or permitted action. For instance, we represent as $(\neg sord, \neg acc)$ the rule prohibiting police officers from accessing personal documents without a search & seizure order.

Axiological rules make the (axiological) obligation or permission of an action dependant on on the action's impact on a value. They are partitioned into two sets: those linking the prohibition of an action to a value demoted by that action; and those linking the permission of an action to a value promoted by that action. We represent axiological prohibitions in the form $(V^x, \neg x)_i$, where V is the value demoted by action x, which is consequently prohibited, and i is the weight of the value. We represent axiological permissions as $(V^x, x)_i$ where V is the promoted value, and x is the consequently permitted action. For instance, let us assume that access to a mobile phone by police officers demotes privacy, which is a reason for prohibiting it, while it promotes public safety, which is consequently a reason to permit it. We can model these rules as $(Priv^{acc}, \neg acc)_{.4}$ and $(Saf^{acc}, acc)_{.6}$.

3 Normative systems

Reasoning with each kind of rules (conceptual, modulation, deontological, or axiological) has different logical properties and therefore requires a different output operator in an architecture of i/o logics (for an introduction to i/o logics see [9] and [12]).

Let L be a standard propositional language with propositional variables and logical connectives: $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\bot$, $\top$. Let $Val = \{V_1^x, V_2^x, ...V_1^y, V_2^y, ...\}$ be a set of values. We say that $N \subseteq G \times G$, where $G \in \{L, Val\}$ is a *normative set* and that each $r \in N$ is a *rule*. For any $A \subseteq G$, $N(G)$ is the image of N under G, that is $N(G) = \{x : (a, x) \in N$, for some $a \in G\}$. We write simply $N(a)$ to abbreviate $N(\{a\})$. To state that x is the output of input a to normative set N, we may write $x \in out_i(N, a)$, or $(a, x) \in out_i(N)$. For any normative set N we define $body(N) = \{a : (a, x) \in N\}$.

Therefore, normative sets contain pairs of propositions or pairs linking a proposition or value to another proposition or value. In order to simplify the exposition, we shall consider actions as propositions (action-propositions), which will be the scope of deontic operators. We shall employ the classical consequence operator Cl. In this paper, it is possible that an action impacts different values. However possible combinations of values will be assessed in the balancing model, so we do not need to consider conjunctions of values or logical inferences among them. Given that we consider values as primitive entities, which are logically independent of each other and that no consequence relations among them are of interest, we shall consider weaker versions of i/o logics, where no consequence operator is applied to the output (of the set of modulation rules). We shall use the following i/o operators:

Definition 3.1 Let N be a normative set, $A \subseteq L$ and $\mathscr{V}$ the set of all maximal consistent sets v in classical propositional logic. Then, we define the following output operators:
(i) *simple minded*: $out_1(N, A) = Cl(N(Cl(A)))$
(ii) *weak*: $out_{1^-}(N, A) = N(Cl(A))$
(iii) *basic*: $out_2(N, A) = \bigcap\{out_1(N, v) : A \subseteq v$, for $v \in \mathscr{V}$ or $v = L$ $\}$
(iv) *weak basic*: $out_{2^-}(N, A) = \bigcap\{out_{1^-}(N, v) : A \subseteq v$, for $v \in \mathscr{V}$ or $v = L\}$

(v) *basic reusable*: $out_4(N, A) = \bigcap\{out_1(N, v) : A \subseteq v$ and $out_1(N, v) \subseteq v$, for $v \in \mathscr{V}$ or $v = L\}$

Definition 3.2 Let N be a normative set and $P \subseteq (L \times L)$ a set of explicit permissions. Then, $(a, x) \in perm_i(P, N)$ iff $(a, x) \in out_i(N \cup Q)$, for some singleton or empty $Q \subseteq P$.

One may combine normative sets N_1 and N_2 and output operators out_i, out_j, 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, that is $out_{i,j}(N_1, N_2, A) = out_i(N_1, out_j(N_2, A) \cup I)$, where $I \in \{A, \emptyset\}$. We call *sequence* a chain of combinations of normative sets.

Definition 3.3 (*Normative System*) Let $A, I \subseteq L$. Let $N_1, ..., N_n, N$ be normative sets and $r \in \{0, 1\}$.
Then $(N_1^{out_1, r_1}, ..., N_n^{out_n, r_n})$, where out_j is the output operator associated to set N_j is a *sequence of normative sets* iff for all N_j, $1 \leq j \leq n$, it holds that $out_j, \ldots out_n(N_j, N_{j+1}, ..., N_n, A) = out_j(N_j, out_{j+1}, ..., out_n(N_{j+1}, ..., N_n, A) \cup I)$, where $N_j \subseteq N$ and $I = A$, if $r_i = 1$, or $I = \emptyset$, if $r_i = 0$. A *normative system* is a class of sequences of normative sets.

We shall write N^{i, r_1} as an abbreviation of N^{out_i, r_1} and $out_{k,l}(N, M, A)$ to abbreviate $out_k, out_l(N, M, A)$.

Our model constructs a particular structure or architecture of normative systems, where the set of conceptual rules (box C) contributes to the determination of which deontological rules and which value assessments are triggered. Following [10] we assume that the set of conceptual rules is governed by a basic reusable output operator and the set of deontological rules is governed by a basic output operator. Their combination is given by the identities: $out_{2,4}(O_d, C, A) = out_2(O_d, out_4(C, A) \cup A)$ and $perm_{2,4}(P_d, C, A) = perm_2(P_d, out_4(C, A) \cup A)$.

From now on, we may write O/P_d or O/P_v for referring to both obligation and permission rules, $O_{d/v}$ and $P_{d/v}$ for both deontological and axiological rules and $O/P_{d/v}$ to include all modalities. The value assessment employs the set of modulation rules and two sets of axiological rule. The set of modulation rules (M) links facts and concepts to the value-impacts of the action in the presence of such fact and concepts. It is governed by a weakened basic output operator. One set of axiological rules (P_v) links each value to the permission of the action that promotes it, and the other (O_v) links each value to the prohibitions of the action that demotes it. Both are governed by the axiological output operator $out_{\succ}$, defined in Section 4.

The combination of these normative sets is given by the following identity:

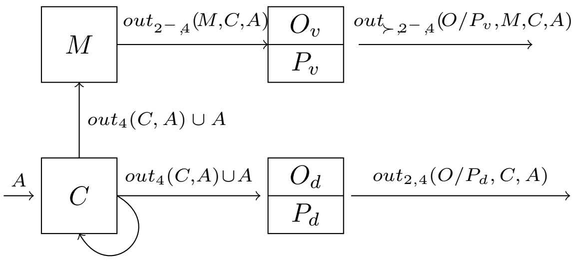
$$out_{\succ,2^-,4}(O/P_v, M, C, A) = out_{\succ}(O/P_v, out_{2^-}(M, out_4(C, A) \cup A))$$

Hence our discussion shall involve the following structures:

$$\langle O/P_d, C\rangle = \{(O_d^{2,0}, C^{4,1}), (P_d^{2,0}, C^{4,1})\}$$
$$\langle O/P_v, M, C\rangle = \{(O_v^{\succ,0}, M^{2^-,0}, C^{4,1}), (P_v^{\succ,0}, M^{2^-,0}, C^{4,1})\}$$

$$\langle O/P_{d/v}, M, C\rangle = \langle O/P_d, C\rangle \cup \langle O/P_v, M, C\rangle$$

The normative system is specified by indicating the rules of each normative set in the corresponding structure. The structure $\langle O/P_{d/v}, M, C\rangle$ of normative systems is represented in the figure below. The arrows indicate the direction of the outputs and inputs of each normative set.



4 Axiological entailment

An axiological entailment presupposes a determination of the comparative moral merits of the choice of performing an action rather than abstaining from it. The action may consist in any behaviour, e.g., having an abortion rather that continuing the pregnancy or accessing an *sms* message, rather than respecting its confidentiality.

The comparison depends on the evaluations expressed by the quantitative indexes of modulation rules (for influence on impact on values) and axiological rules (for weighs of values). For generality's sake we assume that such indexes can take arbitrary numerical assignments within given ranges. These numbers can be restricted to any scales that may be convenient for the chosen domain of application. Here we shall use the positions (0,.2,.4,.6,.8,1) in the examples. What matters is that the numerical assignments reflect some relative importance of the elements at stake, as part of a reasoning with dimensions and magnitudes, and how such assessment of relative importance affects the outputs of the systems and its overall coherence.

4.1 Evaluation of axiological rules

The entailment of axiological rules may involve three kinds of rules –conceptual, modulation, and axiological ones–, so that their evaluation depends on the intensity of factors and the weights of values.

The evaluation model basically compares, for each given action, its impact on the set of values it promotes against its impact on the set of values it demotes, given the constellation of factors, i.e the context in which the action is performed. Two clarifications are of central importance to understand the model here proposed.

First, we only consider the assessment of impact of a single action on values and therefore we only compare the values promoted against the values demoted by that specific action, so that a decision takes place whether that action should or should not be performed on moral grounds. There is no room in this model to compare and decide among different and logically independent actions in terms of their impacts on values. Typically, a claim before a court questions the legality of a particular action and the court must decide whether that action under evaluation should be performed or not (should be forbidden or permitted, should be punished or not be punished). So we keep the same structure regarding its axiological evaluation. We acknowledge that there may be contexts where a judicial decision compares and chooses among alternative courses of action, for instance, between the consumer's right to receive a new product or to have his money back. However we shall leave this kind of value assessment to future work.

Second, we assume that the direction of impact of an action on a value –i.e., whether the action promotes or demotes the value– is invariant, although the extent of the promotion or demotion may be *intensified* or *attenuated* by the presence of factors in the context of performance. By saying that the direction of impact is invariant, we mean that irrespective of how many attenuating factors are taken into account, the impact of an action in the promotion of a particular value never shifts to its demotion. And vice-versa the impact of the action at stake on the demotion of a value never shifts to its promotion.

Let us illustrate the rationality behind the model with an example. Suppose the rules of a condominium forbid people to take the elevator during the pandemics. Suppose now that one inhabitant has a medical emergency. Then one could evaluate whether following the rule would lead to immoral results. The factor "medical emergency" is an intensifier w.r.t the promotion of the value of the patient's health, which would lead to a permission to use the elevator. But now consider that the emergency does not hinder the patient's ability to walk (for instance, it is a toothache) and that she lives in the second floor. So the proportional influence of the set of factors on the promotion of the patient's health may become null or negative, but one would not say that the action of taking the elevator would now demote her health in that particular context. Actually the action still promotes health even in presence of those attenuating factors. But in such cases the proportional impact of the action is so low that it becomes morally irrelevant to legal considerations, that is, it will not play a role in a consideration whether to follow the rule or not. Hence, in the model here proposed attenuating factors only affects the degree of moral impact of the action on a value.

Considering that the direction of impact of the action on a value is invariant, then for a given an action x, the set Val of values may be partitioned into the set of values Val^x_{Dem} which are demoted by the action and a set of values Val^x_{Prom} which are promoted by the action. The relative importance or weight of each value, denoted by w_V, is given by a *weight function* $w : Val \longrightarrow [0, 1]$.

Both features, i.e. the direction of impact of the action on a value and the weight of the value may be directly represented in our architecture by defining the O_v box and P_v box respectively as $O_v = \{(V^x, \neg x)_{w_V} : V \in Val^x_{Dem}\}$ and $P_v = \{(V^x, x)_{w_V} : V \in Val^x_{Prom}\}$.

Let us now move to modulation rules. As noted in Section 2, the extent to which an action promotes or demotes the relevant values is determined by the baseline impact of the action and by the context (the constellation of factors) in which the action takes place. The influence of a factor on the action's impact on a value is given by the *modulation function* $\Delta : L^2 \times Val \longrightarrow [-1, 1]$. We denote by $\Delta^m_V(x)$ the influence of the modulating factor $m \in body(M)$ on the impact of the action $x \in L$ on the value $V^x \in Val$.

Considering that we shall not model the evaluation of sets of different actions, but only the impact of a single action on the promotion against the demotion of given values, we shall omit the reference to the action at stake in the indication of its impact on a value, i.e., we shall indicate such impact with $(V, \neg x)_w \in O_v$ and $(V, x)_w \in P_v$, rather than $(V^x, \neg x)_w \in O_v$ and $(V^x, x)_w \in P_v$.

If the influence of a factor m on a value V is positive ($\Delta^m_V > 0$), m is an *intensifier* of the impact of the action at stake on value V. If the influence is negative ($\Delta^m_V < 0$), then m is an *attenuator* of its impact on V. If there is no influence ($\Delta^m_V = 0$), m is *neutral*.

By a λ-evaluation we mean an evaluation assignment $\lambda_i = [\Delta_i, w_i]$, where Δ_i is a modulation function and w_i is a weight function, and we denote by Λ the set of all λ-evaluations. The proportional influence of a modulating factor m on value V, denoted by ϕ^m_V, is the product of the index of the modulation rule (indicating the intensification or attentuation due to the factor) and of the index of the axiological rule (indicating the weight of the value), that is:

Definition 4.1 (*Proportional influence of a modulating factor on a value*)
Let $(m, V)_{\Delta^m_V} \in M$ and $(V, x)_{w_V} \in O/P_v$. Then: $\phi^m_V = \Delta^m_V \times w_V$

Now we extend the definition of proportional influence to cover the impact of a set of factors B on a set of values W, such an impact being the sum of the proportional influences of each factor.

Definition 4.2 (*Proportional influence of factors on values*) Let $Q = \{(m_1, V_1)_{i_1}, ..., (m_n, V_k)_{i_n}\} \subseteq M$ and $U \subseteq O/P_v$ such that $U = \{(V_1, x)_{j_1}, ..., (V_k, x)_{j_k}\}$. Then, for factors $B = body(Q)$ and values $W = body(U)$ we have:

$$\Phi^B_W = \sum_{1 \leq i \leq n}^{1 \leq j \leq k} \phi^{m_i}_{V_j}$$

4.2 Axiological Output

Given the above definitions, we are able to define the axiological output ($out_\succ$) operator. The idea is to compare the proportional impact of an action on the values it demotes *vis-à-vis* its impact on the values it promotes, considering only those values which are triggered by the input. The set $M(A)$ of the modulating factors involved in the comparison is the subset of $body(M)$, which is triggered by the input $A \subseteq L$, that is, $M(A) = Cl(A) \cap body(M)$. Since the normative system used in our model also includes conceptual rules in the sequence, we have $M(A) = (out_4(C, A) \cup A) \cap body(M)$. In their turn, the sets of values involved in the comparison are those subsets of $body(O_v)$ and of $body(P_v)$, which are triggered by the output of the set of modulation rules. That is, we are going to compare set of demoted values triggered by input A, i.e., $O_v(A) = out_i(M, A) \cap body(O_v)$, against the set of promoted values triggered by input A, i.e., $P_v(A) = out_i(M, A) \cap body(P_v)$. In our architecture, where the sequence includes conceptual rules, we compare $O_v(A) = out_{2^-,4}(M, C, A) \cap body(O_v)$ against $P_v(A) = (out_{2^-,4}(M, C, A) \cap body(P_v)$.

If, for a given constellation of factors, the proportional impact of the action on the *demoted* values is positive and stronger than its proportional impact on the promoted values, then there is an overall axiological prohibition to do it. On the other hand, if the proportional impact of the action on the *promoted* values is positive and stronger that its proportional impact on the demoted values, then there is an axiological explicit permission to do it. Hence, we have the following definition of the value output operator $out_{\succ}$. We may write simply $O/P_v(A)$ to abbreviate $\Phi^{M(A)}_{O/P_v(A)}$.

Definition 4.3 [*Axiological output*] Consider $NS = \langle O/P_v, M \rangle$, $A \subseteq L$ and $x \in L$. Then $x \in out_{\succ}(O_v, M, A)$ iff: (i) $y \in O_v(out(M, A)))$, (ii) $x \in Cl(y)$ and (iii) $0 < O_v(A) > P_v(A)$. The same holds, *mutatis mutandis*, for $out_{\succ}(P_v, M, A)$.

It is worth mentioning that, contrary to all the other output operators discussed so far, the axiological output is defeasible, *i.e.* it does not satisfy the property of Strengthening the Input, according to which if $b \vdash a$, and $(a, x) \in out(N)$, then $(b, x) \in out(N)$ (see example 4.4).

When assessing whether there is convergence of axiological and deontological outputs we need to compare the intensity of the action's impact on each value, relatively to given contexts (constellations of input factors).

A modulating factor may trigger more than one value (directly or indirectly, *i.e.* by detaching other modulating factors) and the impact on a single value may be affected by different modulating factors. Therefore it is interesting to compare modulating factors in terms of the influence of each on the impact on the aggregate of values, as well as to observe how much each value is impacted by the action in a given context.

In order to compare modulating factors, we call the quantity $\Phi^{\{m\}}_{O/P_v(m)}$, where $O/P_v(m) = \{V : V \in out_{2^-}(M, m)\}$, the *strength* of the modulating factor $m \in body(M)$. It represents the sum of the all impacts of the action (on the promoted values or on the demoted values), which are triggered by the modulating factor m. We are going to abbreviate by $m_1 > m_2$ the comparison of strengths of different modulating factors $\Phi^{\{m_1\}}_{O/P_v(m_1)} > \Phi^{\{m_2\}}_{O/P_v(m_2)}$.

In order to compare how much different values are impacted by a given input, we shall use $M(a)$ to denote the set of modulating factors triggered by input a, that is $M(a) = out_4(C, a) \cap body(M)$, and we are going to abbreviate the quantity $\Phi^{M(a)}_{\{V\}}$ by $V(a)$, which represents the extent of the action's impact on a single value V, given input a. So, given $a, b \in L$, the expression $V_1(a) >$

$V_2(b)$ denotes $\Phi^{M(a)}_{\{V_1\}} > \Phi^{M(b)}_{\{V_2\}}$.

Let us illustrate these notations with a hypothetical λ- evaluation, which represents the *Riley vs California* case, where the values of privacy, public safety, and property rights were affected. The US case law before that decision included a rule according to which an officer could access personal property when arresting an individual due to a criminal offense. This rule could be explained by the following considerations on the underlying value impacts: the modulating factor "arrest" intensifies the promotion of public safety (through the action search) so as to outweigh the extent to which the factors property and "personal data" intensify the demotion (through the same action) of property rights and privacy respectively. However, as considered by the court, if the item collected is a mobile phone, then the negative impact on privacy is intensified to the extent that the promotion of public safety is outweighed. This led the court to introduce an exception for searches involving mobile phones.

Example 4.4 [Riley vs California] Consider $NS = \langle O/P_v, M, C\rangle$:
$O_v = \{(Priv, \neg acc)_{.4}, (Pright, \neg acc)_{.4}\}$, $P_v = \{(Saf, acc)_{.6}\}$
$M = \{(\top, Priv)_{.2}, (\top, Pright)_{.0}, (\top, Saf)_{.2}, (dat, Priv)_{.6}, (prop, Pright)_{.4},$
$(arrest, Saf)_{.8}, (mob, Priv)_1\}$, $C = \{(mob, data), (mob, prop)\}$
We have that $Saf(arrest) = 0.6$, $Pright(prop) = 0.16$, $Priv(data) = 0.32$ and the factor mobile played a strong intensifying role with $Priv(mob) = 0.72$. The strength of factors each factor is $mob = 0.88$, $arrest = 0.6$, $dat = 0.32$ and $prop = 0.16$. So, we have $mob > arrest > dat > prop$ and, comparing the values, it holds that $Saf(arrest) > Pright(prop) + Priv(dat)$, but $Priv(mob) + Pright(mob) > Saf(arrest)$.

Hence, the balancing above explains the shift in the U.S case law given the factor "mobile phone", as we have that $acc \in out_{\succ}(P_v, M, C, \{arrest, prop, data\})$, but it also holds that $\neg acc \in out_{\succ}(O_v, M, C, \{arrest, mob\})$. That is, it is morally admissible for the police to access property items and personal data in an arrest, but it is immoral to access the content of a mobile phone, for the impact on privacy, in that case, is severely intensified (a mobile phone is conceptually both property and data).

Based on the strength of the impacts on the values triggered by an input, we define the proportional impact of an entailed axiological rule as the difference between the values promoted (demoted) and demoted (promoted) in the entailment.

Definition 4.5 *(Proportional impact of a rule)* Consider a normative system $NS = \langle O/P_v, M, C\rangle$, and $(a, x) \in out_{\succ}(O_v, M, C)$. Then, $\sigma(a, x) = O_v(a) - P_v(a)$ is the proportional impact of the rule (a, x). The same holds, *mutatis mutandis, for* $(a, x) \in out_{\succ}(P_v, M, C)$.

In Example 4.4, the proportional impact of prohibiting access to the content of a mobile phone in an arrest is $\sigma(mob \wedge arrest, \neg acc) = 0.28$.

5 Normative theories

A normative system is the object of assertions by jurists (legal doctrine) who describe the systems through normative propositions, i.e., statements that certain obligations and permissions hold given certain factors, according to a normative system.

Normative propositions, while being descriptive of a given normative system (as viewed by the interpreter), also reflect the evaluative aspects of the described system, namely, the ascription of intensities of influence (to modulation rules) or the ascription of weights of values (to axiological rules). Such λ-evaluations contribute to determine the axiological obligations/permissions delivered by the system, and consequently, what normative propositions would be true about it.

Definition 5.1 Let $NS = \langle O/P_{d/v}, M, C\rangle$ be a normative system and $b, x \in L$. The for a given λ evaluation:
$NS \models^{\lambda} \mathbb{O}_d(x/b)$ iff $x \in\ out_{2,4}(O_d, C, b)$
$NS \models^{\lambda} \mathbb{P}^{-}{}_d(x/b)$ iff $\neg x \notin\ out_{2,4}(O_d, C, b)$
$NS \models^{\lambda} \mathbb{P}^{+}{}_d(x/b)$ iff $x \in\ perm_{2,4}(O_d, P_d, C, b)$
$NS \models^{\lambda} \mathbb{O}_v(x/b)$ iff $x \in out_{>,2^-,4}(O_v, M, C, b)$
$NS \models^{\lambda} \mathbb{P}^{-}{}_v(x/b)$ iff $\neg x \notin out_{>,2^-,4}(O_v, M, C, b)$
$NS \models^{\lambda} \mathbb{P}^{+}{}_v(x/b)$ iff $x \in out_{>,2^-,4}(P_v, M, C, b)$

Each normative proposition describes an entailed deontological or axiological rule, with the exception of negative permissive propositions, which describe the non-derivability of such a rule. Thus, following Alchourrón [1], we distinguish a negative sense of permission $\mathbb{P}^{-}{}_{d/v}(x/b)$, as the absence of prohibition, from a positive sense of permission as an entailed deontological or axiological permission $\mathbb{P}^{+}{}_{d/v}(x/b)$.

A normative theory Th^{λ}_{NS} about a normative system NS is the set of all normative propositions describing the rules entailed by that normative system based on the λ-evaluation, that is on given modulation and weight functions: $Th^{\lambda}_{NS} = \{\alpha : NS \models^{\lambda} \alpha\}$. We say that a normative system leads to a conflict, relatively to a certain input factors when, given that input, the systems delivers the prohibition and the permission of the same action. We distinguish conflicts of normative propositions according to the kind of rules which contribute to produce the conflict:

Definition 5.2 (*Consistency, Coherence and Stability of normative theories*) For any given $b \in L$, a normative theory is:
b-inconsistent iff $\bot \in out_2(O/P_d, b)$; b-incoherent iff $\bot \in out_{2,4}(O/P_d, C, b)$;
b-λ-unstable iff there is $x \in L$ for which $\{\mathbb{O}_v(\neg x/b), \mathbb{P}_d(x/b)\} \subseteq Th^{\lambda}_{NS}$ or $\{\mathbb{O}_d(\neg x/b), \mathbb{P}_v(x/b)\} \subseteq Th^{\lambda}_{NS}$

In other words, inconsistency captures cases in which deontological rules directly deliver incompatible conclusions, proper incoherence the case in which the conflict of deontological rules is triggered by a conceptual classification, and proper instability the case in which deontological rules are in conflict with axiological rules. We also may say that a normative theory is strongly stable, relatively to an input, if the corresponding deontological normative propositions are matched by corresponding axiological proposition, and that it is weakly stable, if the deontological propositions are not conflicted by axiological propositions.

We propose here an interpretation of Radbruch's formula, based on the concept of "proportional impact" of a rule, as the key to define different entailment relations and, accordingly, different legal theories.

Definition 5.3 (*Negative Radbruch's Formula*) Let $NS = \langle O/P_{d/v}, M, C\rangle$ be a normative system, $b, x \in L$ and λ an evaluation, then:
(i) $NS \models^{\lambda}_{rad^-} \mathbb{P}^+(x/b)$ iff $NS \models^{\lambda} \mathbb{P}^+_d(x/b)$ and it is not the case that $NS \models^{\lambda} \mathbb{O}_v(\neg x/b)$ and $\sigma(b, x) \geq r$, where r is a treshold index;
(ii) $NS \models^{\lambda}_{rad^-} \mathbb{O}(x/b)$ iff $NS \models^{\lambda} \mathbb{O}_d(x/b)$ and it is not the case that $NS \models^{\lambda} \mathbb{P}^+_v(\neg x/b)$ and $\sigma(b, x) \geq r$.

According to Definition 5.3, morality only has a censorial role: it produce no legal conclusions and only excludes the application of highly immoral deontological rules.

Definition 5.4 (*Positive Radbruch's Formula*) Let $NS = \langle O/P_{d/v}, M, C\rangle$ be a normative system, $b, x \in L$ and λ an evaluation, then:
(i) $NS \models^{\lambda}_{rad^+} \mathbb{P}^{+/-}(x/b)$ iff $NS \models^{\lambda} \mathbb{P}^{+/-}_d(x/b)$ and it is not the case that both $NS \models^{\lambda} \mathbb{O}_v(\neg x/b)$ and $\sigma(b, x) \geq r$; otherwise $NS \models^{\lambda} \mathbb{O}(\neg x/b)$
(ii) $NS \models^{\lambda}_{rad^+} \mathbb{O}(x/b)$ iff $NS \models^{\lambda} \mathbb{O}_d(x/b)$ and it is not the case that both $NS \models^{\lambda} \mathbb{P}^+_v(\neg x/b)$ and $\sigma(b, x) \geq r$; otherwise $NS \models^{\lambda} \mathbb{P}^+(\neg x/b)$

According to Definition 5.4, morality has both a censorial role and a generative one, delivering outputs with high moral merit (proportional impact above threshold).

Definition 5.5 (*Dual Radbruch's Formula*) Let $NS = \langle O/P_{d/v}, M, C\rangle$ be a normative system, $b, x \in L$ and λ an evaluation, then:
(i) $NS \models^{\lambda}_{dual} \mathbb{P}^{+/-}(x/b)$ iff $NS \models^{\lambda} \mathbb{P}^{+/-}{}_v(x/b)$ and it is not the case that both $NS \models^{\lambda} \mathbb{O}_d(\neg x/b)$, and $\sigma(b, x) \leq r$; otherwise $NS \models^{\lambda} \mathbb{O}(\neg x/b)$
(ii)$NS \models^{\lambda}_{dual} \mathbb{O}(x/b)$ iff $NS \models^{\lambda} \mathbb{O}_v(x/b)$ and it is not the case that both $NS \models^{\lambda} \mathbb{P}^+{}_d(\neg x/b)$ and $\sigma(b, x) \leq r$; otherwise $NS \models^{\lambda} \mathbb{P}^+(\neg x/b)$

According to Definition 5.5, morality has both a censorial role and a generative one. The difference from Definition 5.4 lies in those cases where axiological outputs are not conflicted by deontological rules. By the Dual Radbruch's formula all such axiological outputs are delivered by the legal system, while in the Positive Radbruch Formula an axiological output is only delivered when it exceeds the moral threshold.

Our definitions of the Radruch formulas also cover the limit cases when the threshold is null ($r = 0$) or infinite ($r = \infty$). This allows us to capture eight different legal theories, which differs with respect to the specific question whether external considerations of morality may generate valid law.

Definition 5.6 Let $NS = \langle O/P_{d/v}, M, C\rangle$ be a normative system, $b, x \in L$, λ an evaluation and i a given threshold in a Radbruch's formula. Then:

- Closed Positivism [2]: $NS \models^{\lambda}_{pos} \mathbb{O}/\mathbb{P}(x/b)$ iff $NS \models^{\lambda}_{rad} \mathbb{O}/\mathbb{P}(x/b)$ and $r = \infty$
- Open Positivism: $NS \models^{\lambda}_{opos} \mathbb{O}/\mathbb{P}(x/b)$ iff $NS \models^{\lambda}_{dual} \mathbb{O}/\mathbb{P}(x/b)$ and $r = \infty$
- Strong Censorial Non-Positivism: $NS \models^{\lambda}_{snnp} \mathbb{O}/\mathbb{P}(x/b)$ iff $NS \models^{\lambda}_{rad^-} \mathbb{O}/\mathbb{P}(x/b)$ and $r = 0$
- Weak Censorial Non-Positivism: $NS \models^{\lambda}_{wnnp} \mathbb{O}/\mathbb{P}(x/b)$ iff $NS \models^{\lambda}_{rad^-} \mathbb{O}/\mathbb{P}(x/b)$ and $0 < r < \infty$
- Strong Generative Non-Positivism: $NS \models^{\lambda}_{spnp} \mathbb{O}/\mathbb{P}(x/b)$ iff $NS \models^{\lambda}_{rad^+} \mathbb{O}/\mathbb{P}(x/b)$ and $r = 0$
- Weak Generative Non-Positivism: $NS \models^{\lambda}_{wpnp} \mathbb{O}/\mathbb{P}(x/b)$ iff $NS \models^{\lambda}_{rad^+} \mathbb{O}/\mathbb{P}(x/b)$ and $0 < r < \infty$
- Absolute Natural Law: $NS \models^{\lambda}_{anl} \mathbb{O}/\mathbb{P}(x/b)$ iff $NS \models^{\lambda}_{dual} \mathbb{O}/\mathbb{P}(x/b)$ and $r = 0$
- Relative Natural Law: $NS \models^{\lambda}_{rnl} \mathbb{O}/\mathbb{P}(x/b)$ iff $NS \models^{\lambda}_{dual} \mathbb{O}/\mathbb{P}(x/b)$ and $0 < r < \infty$

For *Closed Positivism* only deontological outputs are delivered, while axiological outputs are irrelevant to legal validity. For *Open Positivism* all deontological outputs are delivered together with the axiological outputs that are consistent (not conflicting) with them. For *Strong Negative Natural Law*, only those deontological outputs are valid, which are consistent with all axiological outputs. For *Weak Negative Natural law*, the deontological outputs are delivered, which are not inconsistent with those highly ranked axiological outputs above the assumed threshold. For *Strong Positive Natural law*, all axiological outputs are delivered plus those deontological outputs that are consistent with them. For *Weak Positive Natural Law*, those axiological outputs with high proportional impact (above the treshold) are delivered together with those deontological outputs which are consistent with them. For *Absolute Natural Law*, only axiological outputs are delivered. For *Relative Natural Law*, axiological outputs of high proportional impact are delivered independently of consistency with deontological outputs, while axiological outputs of lesser impact are delivered only if consistent with delivered deontological outputs. One could say, in a theory resembling Finnis' [6], that Relative Natural Law would contend that matters with low moral significance (e.g. coordination problems) could be left to discretionary choices by authorities, while sensitive matters should be ruled by moral reasoning.

In the table below we present the output of the normative systems for each legal theory concerning the *Riley* case and the corresponding theoretical explanation for the court's decision to prohibit access to the mobile phone. Notice that the positively enacted law provides the deontological output that ${\mathbb{P}^+}_d(acc/mob \wedge arrest)$, while the axiological output is $\mathbb{O}_v(\neg acc/mob \wedge arrest)$ with a proportional impact $\sigma(mob \wedge arrest, \neg acc) = 0.28$, according to the

[2] using the rad^- or the rad^+ entailment relation results in the same positivist theory of validity.

Riley v California	Output of the normative system	Explanation of court decision
Closed Positivism	Permitted	Change the Law
Open Positivism	Permitted	Change the Law
Strong Censorial Non-Positivism	Gap	Fill
Weak Censorial Non-Positivism	Gap/Permitted	Fill/Change the Law
Strong Generative Non-Positivism	Forbidden	Apply the Law
Weak Generative Non-Positivism	Permitted / Forbidden	Apply/Change the Law
Absolute Natural Law	Forbidden	Apply the Law
Relatve Natural Law	Forbidden	Apply/Change the Law

assumed λ-evaluation. The normative propositions describing the content of the normative system would be either a positive permission, a prohibition or a negative permission (a gap). According to these theories the decision of the U.S court –forbidding access to the content of the mobile phone– would have different explanations: that the court changed the existing law (*contra legem*), that it applied the existing law (*secundum legem*), or that it filled a gap by discretion creating new law (*extra legem*).

With respect to the weak versions of non-postivism, the outcome of the normative systems and the corresponding explanations would depend on whether the theory assumes a Radbruch's threshold above or below 0.28. Suppose the theory assumed a threshold $r = 0.6$. Then the weak non-positivist theory would maintain that it is permitted to access the content of a mobile phone in an arrest, since the reached level of immorality is below the threshold. But now suppose that that positive law authorized the search of an individual's mobile phone independently of any arrest. Then the axiological output would be $\mathbb{O}_v(\neg acc/mob)$ with a proportional impact $\sigma = 0.76$ (thus above the 0.6 threshold). Therefore the final outcome, for accessing the content of the mobile phone independently of an arrest would be either a gap (weak non-positivism) or a prohibition to access (strong non-positivism).

6 Final Remarks

By combining an architecture of i/o logics and a model of balancing values, we have proposed a formal concept of normative system, where obligations and permissions may be assessed in terms of their impact on the promotion or demotion of moral values. Based on this concept and on three interpretations of the so-called Radbruch's formula, we have formally defined eight different conceptions of the connection between law and morality. The above analysis assumes that axiological consideration are external to the positively enacted law, pertaining to political morality. However, our approach is also compatible with the assumption that axiological considerations are internal to the positively enacted law, as legal principle o fundamental rights, in particular those enshrined in a Constitution. In future investigations, following the latter approach, we may define corresponding versions of Constitutionalism from classical negative (censorial) constitutionalism to different generative forms of neo-constitutional moralism and principialism.

The framework here proposed brings together two parallel lines of research in AI & law: on the one hand the study of the role of values in case-based legal argumentation ([14] and [15]), and on the other hand the study of statutory interpretation as the dynamical modification of combined normative sets, including conceptual qualification, conditional rules and values ([3], [10] and [11]). One of the difficulties in the last approach is how to set up and formalize criteria for the choice between alternative normative systems that satisfy a revision function. This paper offers a conceptual and formal basis for the balancing of values that may be used as criteria both to trigger and to choose between possible results of revisions of normative systems. The modelling of constructive legal interpretation by revision functions based on the framework here proposed will be left to future work, as well as the effort to characterize the axiological output operator here advanced.

References

[1] Alchourrón, C., *Logic of norms and logic of normative propositions*, Logique et analyse **12** (1969), pp. 242–68.

[2] Alexy, R., "The Argument from Injustice. A Reply to legal positivism," Oxford University Press, 2003.

[3] Boella, G. and L. van der Torre, *A logical architecture of a normative system*, in: *Deontic Logic and Artificial Normative Systems - DEON 2006* (2006), pp. 24–35.

[4] Dancy, J., "Ethics Without Principles," Oxford University Press, 2004.

[5] Dworkin, R. M., "Law's Empire," Kermode, 1986.

[6] Finnis, J. M., "Natural Law and Natural Rights," Clarendon, 1980.

[7] Grossi, D., J.-C. Meyer and F. Dignum, *The many faces of counts-as: A formal analysis of constitutive rules*, Journal of Applied Logic **6** (2008), pp. 192–217.

[8] Hart, H. L. A., *Positivism and the separation of law and morals*, in: *Essays in Jurisprudence and Philosophy*, Clarendon, 1983 pp. 49–87, (1st ed. 1958.).

[9] Makinson, D. and L. van der Torre, *Input/output logics*, Journal of Philosophical Logic **29** (2000), pp. 383–408.

[10] Maranhão, J., *A logical architecture for dynamic legal interpretation*, in: *Proceedings of the Eight International Conference on AI and Law ICAIL '17,*, ACM Press, 2017 pp. 129–38.

[11] Maranhão, J. and E. G. de Souza, *Contraction of combined normative sets*, in: *Deontic Logic and Normative Systems: 14th International Conference, DEON 2018*, Springer, 2018 pp. 247–261.

[12] Parent, X. and L. van der Torre, *Input/output logic*, in: J. Horty, D. Gabbay, X. Parent, R. van der Meyden and L. van der Torre, editors, *Handbook of Deontic Logic and Normative Systems, Volume 1*, College Publications, 2013 pp. 499–544.

[13] Radbruch, G., *Statutory lawlessness and supra-statutory law*, Oxford Journal of Legal Studies **6** (2006), pp. 1–11, (1st ed. 1946.).

[14] Sartor, G., *The logic of proportionality: Reasoning with non-numerical magnitudes*, German Law Journal (2013), pp. 1419–57.

[15] Sartor, G., *Consistency in balancing: from value assessments to factor-based rules*, in: D. Duarte and S. Sampaio, editors, *Proportionality in Law: An Analytical Perspective*, Springer, 2018 pp. 121–36.