Permissions and Undercutters

Guido Boella
Dipartimento di Informatica
Università di Torino
Italy
guido@di.unito.it

Leendert van der Torre
SEN3
CWI Amsterdam
The Netherlands
torre@cwi.nl

Abstract

Input/output logics have been proposed to formalize rule based inference. They are mainly inspired by deontic logic, or the logic of obligations, although it has also been shown that they generalize for example Reiter's normal default logic. A recent extension of the input/output logics formalizes various notions of permission. We are interested in the question whether input/output logics can be used to formalize rule based inferences as they occur in for example belief revision, default reasoning, argumentation, causal reasoning, *et cetera*, and in particular whether permissions correspond to notions in such inferences. In this paper, we discuss permissions and their role in deontic logic, and relate it to undercutters in argumentation theory.

1 Introduction

It is well known that there are relations between various approaches to reasoning with conditionals. For example, Hansson's original normative conditional $O(a|b)$ [Hansson, 1969] – the minimal b worlds satisfy a – has been related to counterfactual reasoning by Lewis [1974], it has been used as a default conditional by many authors in the tradition of Shoham [Shoham, 1988; Boutilier, 1994a], which has itself been related to belief update and revision [Boutilier, 1994c], it has been used as a conditional for ideality statements by Boutilier [1994b], *etcetera*. Makinson's study of five faces of minimality [Makinson, 1993] also shows subtle distinctions between various kinds of reasoning based on minimality.

The role of the dual operator of permission and the possible relation to dual operators in other rule based inferences has received less attention. Originally, von Wright [1951] – the founding father of deontic logic – introduced his deontic logic as a modal system due to a correspondence of necessity and possibility on the one hand, and obligation and permission on the other hand. However, he soon criticized this approach, arguing that the duality reading of permissions is limited to a notion of so-called negative permissions. Other notions of so-called positive permissions have been studied. Whether these notions correspond to notions studied in other areas seems not to have been discussed in the literature.

In this paper, we are interested in the question whether permissions can be related to undercutters in argumentation theory, when obligations are related to arguments and rebutters. Consider the following story discussed by Pollock [1987]. “For example, something's looking red to me may justify me in believing that it is red, but if I subsequently learn that the object is illuminated by red lights, and I know that that can make things look red when they are not, then I cease to be justified in believing that the object is red.” Pollock formalizes this example with an argument that the object is red because it looks red, and an undercutting defeater that one should not believe that the object is red if it is illuminated by red light. The undercutting argument thus behaves like a permission, in the sense that it refers to the absence of being red, without implying that the object is not red.

In this paper we use Makinson and van der Torre's input/output logic [2000; 2001; To appear] to discuss the relation between permissions and undercutters. We focus on permissions, and we spend relatively little space on how they can be related to undercutters, because input/output logics only recently emerged and we expect that most researchers in the workshop audience are not familiar with them, and neither with the literature on permission. In particular, we discuss the philosophical perspective on various notions of permission, and the technical development of a formalization of permissions within the context of input/output logics, and we give a preliminary definition of rebutters and undercutters in this framework.

The layout of this paper is as follows. In Section 2 we survey the philosophical logic literature on permissions. In Section 3 we survey Makinson and van der Torre's permissions from an input/output perspective. In Section 4 we introduce our own notion of permissions from an input/output perspective. In Section 5 we discuss several arguments in this setting. Related work and Concluding remarks close the paper.

2 Philosophical considerations

The central problem in the treatment of permissions in deontic logic is whether permission is an autonomous normative category. The question arises because of Von Wright's original interdefinability of deontic operators of permission and obligation by $P(q)$ iff $\neg O(\neg q)$. Is permission only the mere absence of obligation or something which can be positively expressed by a norm?

A negative answer to the latter question has been given by Ross: "I know of no permissive legal rule which is not logically an exemption modifying some prohibition, and interpretable as the negation of an obligation" [Ross, 1968, p. 122]. This conclusion is based on the view that without the context of an obligation a permission is not useful: "telling me what i am permitted to do provides no guide to conduct unless the permission is taken as an exception to a norm of obligation (which may be understood as the general maxim that what is not permitted is prohibited)".

Ross' conclusion can be further motivated by Lewis' "master and slave" game [Lewis, 1979]. In this game, a master classifies the slave's actions into two categories or spheres: the sphere of prohibited actions and the sphere of permitted (i.e., not forbidden) actions or "the sphere of permissibility". One conclusion from this game is that the notion of permission is not enough to build a normative system, because only obligation norms (and prohibition, defined as obligation to the contrary) are able to classify the actions. A similar dynamic perspective has more recently been discussed by Alchourrón [1993].

However, according to von Wright, there are two types of permissions, *weak permissions* are mere absence of obligations, while *strong permissions* (also called explicit permissions) are the content of so called permissive norms. In his earliest paper [von Wright, 1959] the distinction is motivated on the basis of the distinction between weak and strong negation, because weak norms consist of the mere absence of the contrary norms. In his later work [von Wright, 1963], Von Wright argues that the strong permissions provide an answer to the need of disciplining new kinds of behaviors, filling gaps in law and needed to express competence laws. Competence here stands for the ability to change the normative status of actions.

Legal scholars agree with a distinction between two kinds of permissions. For example, Bulygin argues that "the role played by permissive norms is not exhausted by derogation of former prohibition: an act of permitting an action which has not been hitherto prohibited is not at all pointless as has been suggested by those who deny the importance of permissive norms" [Bulygin, 1986, p.213]. However, they challenge the justifications provided by von Wright. For example, Ross argues that "the constitutional guarantee of certain freedoms [...] is a restriction of the power of the legislator, a disability which corresponds to an immunity on the part of the citizen" ([Ross, 1968, p.125]), rather then an imperative or permissive norm on competence. While gaps are better defined as constitutional rights which are not protected by laws so that they do not find a full accomplishment.

The legal scholars individuate the main role of permissive norms in specifying exceptions to obligations. Bobbio explains that "the difference between weak and strong permission becomes clear when we think about the function of permissive norms. Permissive norms are subsidiary norms: subsidiary in that their existence presupposes the existence of imperative norms [...] a permissive norms is necessary when we have to repeal a preceding imperative norm or to derogate to it. That is to abolish a part of it (that in this case it is not necessary preexisting because a law itself may prescribe a limit to its own extension)" [Bobbio, 1980, p.891-892].[1]

Bulygin gives a counterargument to the master-slave argument by arguing that this game lacks the idea that a normative system does not consist of a single authority which enacts norms. Alchourrón and Makinson support the importance of hierarchies: "when we consider the regulations in legal or administrative code, we can often discern some kind of hierarchy among them. Some are regarded as more basic or fundamental than others", [Alchourron and Makinson, 1981, p.125]. Bulygin argues that when a normative system is composed by many authorities which are linked by hierarchical relations, then permissive norms are not superfluous even if there is no corresponding prohibition. This is due to the fact that a normative system has a dynamic character, in the sense that norms are added to the system one after the other and this operation is performed by different authorities at different levels of the hierarchy. It is only in this interpretation of a normative system that permissive norms show their full relevance: "only in a dynamic perspective of a hierarchically structured normative system (with a plurality of norm authorities belonging to different levels) that changes in the source of time as a result of different normative acts carried out by norm authorities where the concept of a permissive norm becomes really fruitful" [Bulygin, 1986, p.216].

Once we consider a plurality of authorities we need to cope with conflicts: the law uses *meta-norms*. The meta-norms of the system ascribe to each level of authority an area of competence and prescribe that the system must respect normative principles like "lex superior derogat inferiori" ("norms have the function of preventing - inhibit, preclude - the creation of imperative norms by subordinated sources of law", [Guastini, 1998] p.29) or "lex posterior derogat priori" (the function of abrogating preexisting imperative norms or to derogate to them).

Such principles are not logical principles, but rather meta-norms on the *validity* of norms, so they must be kept distinct from the core of the logic: different legal systems can be characterized by different meta-norms.

[1] Authors' translation from Italian text.

3 Permissions and input/output logic

Our formal discussion is based on a input/output logic [Makinson and van der Torre, 2000].

To be more accurate, input/output logic has its source in a tension between the philosophy of norms and formal work of deontic logicians. Philosophically, it is widely accepted that a distinction may be drawn between norms on the one hand, and declarative statements on the other. Declarative statements may bear truth-values, in other words are capable of being true or false; but norms are items of another kind. They may be respected (or not), and may also be assessed from the standpoint of other norms, for example when a legal norm is judged from a moral point of view (or vice versa). But it makes no sense to describe norms as true or as false.

Input/output logic not treat conditional norms as bearing truth-values. They are not embedded in compound formulae using truth-functional connectives. To avoid all confusion, they are not even treated as formulae, but simply as ordered pairs (a, x) of purely boolean (or eventually first-order) formulae. Technically, a normative code is seen as a set G of conditional norms, i.e. a set of such ordered pairs (a, x). For each such pair, the body a is thought of as an input, representing some condition or situation, and the head x is thought of as an output, representing what the norm tells us to be desirable, obligatory or whatever in that situation. The task of logic is seen as a modest one. It is not to create or determine a distinguished set of norms, but rather to prepare information before it goes in as input to such a set G, to unpack output as it emerges and, if needed, coordinate the two in certain ways. A set G of conditional norms is thus seen as a transformation device, and the task of logic is to act as its 'secretarial assistant'.

In this paper we only consider the proof theory of input/output logic, not its semantics. We start with obligations. The distinguishing feature of obligations defined in input/output logic is that they are defined relative to a set of generators which may be interpreted as the set of norms. In most other logics of obligations, also called deontic logics, the set of norms (or normative system) is assumed to be implicit.

Definition 1 (Obligations) *Let L be a base logic with $\top$ a tautology, and let G be a set of ordered pairs of L (called the generators). A generator (a, x) is read as 'if input a then output obligatory x'. An input/output logic out is a closure operation on $G \cup \{(\top, \top)\}$ under replacement of logical equivalents, the rules SI, WO and AND, and a subset of OR, CT and ID.*

$$SI \quad \frac{(a,x)}{(a \wedge b, x)} \qquad WO \quad \frac{(a,x)}{(a, x \vee y)} \qquad AND \quad \frac{(a,x),(a,y)}{(a, x \wedge y)}$$

$$OR \quad \frac{(a,x),(b,x)}{(a \vee b, x)} \qquad CT \quad \frac{(a,x),(a \wedge x, y)}{(a,y)} \qquad ID \quad \frac{}{(a,a)}$$

Example 1 *Given the set of obligations $G = \{(a, x), (a, y), (x, z)\}$ a possible output is $\{(a \wedge b, x), (a \wedge x, z), (a, x \vee y), (a, a \vee x), (a, x \wedge y)\}$ using rules SI, WO and AND. Using also the CT rule, the output includes $\{(a, z)\}$.*

Permissions are more ambiguous than obligations, and different notions have been defined [Makinson and van der Torre, To appear]. Makinson and van der Torre distinguish three notions of permission. *negperm* is the negation of an obligation, it corresponds to what is called weak permission. *statperm* guides the citizen in the deontic assessment of specific actions, and behaves like a weakened obligation: given what is obligatory and what is strongly permitted the actual permissions of an agent are computed. They are called weakened obligations since $statperm(P, G) \subseteq out(P \cup G)$, see [Makinson and van der Torre, To appear] for details. *dynperm* guides the legislator by describing the limits on what may be prohibited without violating static permissions, which is called prohibition immunity.

Definition 2 (Permissions) *Let G and P be two sets of generators, where P stands for permissive norms, and let out be an input/output logic.*

- *$(a, x) \in negperm(G)$ iff $(a, \neg x) \notin out(G)$;*
- *$(a, x) \in statperm(P, G)$ iff $(a, x) \in out(G \cup Q)$ for some singleton or empty $Q \subseteq P$;*
- *$(a, x) \in dynperm(P, G)$ iff $(c, \neg z) \in out(G \cup \{(a, \neg x)\})$ for some pair $(c, z) \in statperm(P, G)$ with c consistent.*

Example 2 *It is obligatory to make homework, but if one does homework he is permitted to watch the television $G = \{(\top, h)\}, P = \{(h, t)\}$. Then $(\top, h) \in negperm(G)$, since what is obligatory is permitted and $(a, b) \in negperm(G)$ since given a there is no restriction about b. Moreover, $(h, t) \in statperm(P, G)$ since this is explicitly permitted and $(a, t) \in dynperm(P, G)$: $(a \wedge h, \neg t) \in out(G \cup \{(a, \neg t)\})$ for some pair $(a \wedge h, t) \in statperm(P, G)$.*

The main problem of reasoning with obligations and permissions is the question how to deal with violations and obligations resulting from violations, known as contrary-to-duty reasoning. It has been discussed in the context of the notorious contrary-to-duty paradoxes such as Chisholm's and Forrester's paradox. It has led to the use of constraints in input/output logics [Makinson and van der Torre, 2001].

The strategy is to adapt a technique that is well known in the logic of belief change - cut back the set of norms to just below the threshold of making the current situation contrary-to-duty. In effect, input/output logic carries out a contraction on the set G of generators.

In input/output logics under constraints, a set of generators and an input have as output a set of set of propositions. We can infer a set of propositions by for example taking the join (credulous) or meet (sceptical), or something more complicated. In this paper we only consider the input/output constraints.

Definition 3 (Constraints) *Let G be a set of generators and out be an input/output logic. Moreover, we write $x \in out(G, a)$ iff $(a, x) \in out(G)$. We define:*

- *$maxfamily(G, a)$ is the set of $\subseteq$-maximal subsets G' of G such that $out(G', a) \cup \{a\}$ is consistent.*

- *outfamily*(G, a) *is the output under the elements of maxfamily, i.e.* $\{out(G', a) \mid G' \in$ *maxfamily*$(G, a)\}$.
- $(a, x) \in out_\cup(G)$ *iff* $x \in \cup$*outfamily*(G, a)
  $(a, x) \in out_\cap(G)$ *iff* $x \in \cap$*outfamily*(G, a)

In case of contrary to duty obligations, the input represents something which is inalterably true, and an agent has to ask himself which obligations (output) this input gives rise to: even if the input should have not come true, an agent has to "make the best out of the sad circumstances" [Hansson, 1969]. Makinson and van der Torre [2001] illustrate constraints with the following example.

Example 3 *Multiple level of violation may be analyzed [...]. For example, put* $G = \{(\top, \neg a), (a, x), (a \wedge \neg x, y)\}$ *where [a is read as 'you break your promise', x as 'you apologize' and] y 'you are ashamed'. Consider the input* $a \wedge \neg x$. *Then* $out(G, a \wedge \neg x) = Cn(\neg a, x, y)$, *which is inconsistent with input* $a \wedge \neg x$, *so that maxfamily*$(G, a \wedge \neg x) = \{(a \wedge \neg x, y)\}$ *and finally outfamily*$(G, a \wedge \neg x) = \{Cn(y)\}$.

4 Permissions and priorities

Makinson and van der Torre's discussion on permissions from an input/output perspective does not discuss permissions as exceptions and hierarchies of authorities, which have been identified as crucial elements in the philosophical literature. This is also the case when permissions under constraints are formalized by replacing in Definition 2 each occurrence of *out* by $out_\cup$ or $out_\cap$ [Boella and van der Torre, 2003d]. Most exceptions in the criminal code can be understood as such permissions, e.g., consider "it is forbidden to kill ($(\top, \neg k) \in G$), but it is permitted to kill in self-defense ($(s, k) \in P$), unless a policeman is killed ($(s \wedge p, \neg k) \in G$)". In input/output logic with constraints, these norms still imply the prohibition to kill in case of self-defense ($(s, \neg k) \in out_{\cup/\cap}(G)$), because *maxfamily* and *outfamily* do not take permissions into account. To formalize such permissions, we have to introduce a way for permissions to block or override obligations. Note that again it is the dynamics of a normative system that highlights the importance of permissions-as-exceptions. It is possible that a permission is enacted in the same norm as the obligation it is an exception to, but it is also possible to modify the normative system by adding new permissions after some time, as well as to introduce obligations which are exceptions to some previous permission.

Without permissions, a conflict can be defined in input/output logic as a case in which *outfamily* contains more than one element. Conflict resolution has been studied in defeasible deontic logic. It can be formalized by an ordering on the powerset of generators, such that a *preffamily* selects from *maxfamily* only the preferred elements. An extension needed here is the distinction between generator pointers and the generators themselves, because the same generator may occur several times in the ordering. In fact, the same generator can be the object of norms enacted by different authorities: however, all these instances of the generator may have different priorities. So, we consider each norm, i.e., each instance of a generator, as a different generator pointer. As usual we assume that the ordering is at least a partial pre-order, i.e. antisymmetric and transitive, and that it contains the subset-ordering.

With permissions, there are many ways in which the system can be extended. In this paper we use the definition we give in [Boella and van der Torre, 2003d].

Definition 4 (Permissions as exceptions) *Let G and P be disjoint sets of generators pointers, V a function that associates with every generator pointer a generator, and* $\leq$ *a partial pre-order on the powerset of* $G \cup P$ *that contains the subset-ordering. We read* $A \leq B$ *as B is preferred to A.*

- *maxfamily*(G, P, V, a) *is the set of* $\subseteq$*-maximal* $G' \cup P'$ *such that* $G' \subseteq G$, $P' \subseteq P$ *and* $out(V(G') \cup V(Q), a) \cup \{a\}$ *is consistent for every singleton or empty* $Q \subseteq P'$.
- *preffamily*$(G, P, V, \leq, a)$ *is the set of* $\leq$ *maximal elements of maxfamily*(G, P, V, a).
- *outfamily*$(G, P, V, \leq, a)$ *is the output related to preffamily, i.e.* $\{out(V(G'), a) \mid G' \cup P' \in$ *preffamily*$(G, P, V, \leq, a), G' \subseteq G, P' \subseteq P\}$.
- *statpermfamily*$(G, P, V, \leq, a)$ *is defined analogously for permissions, i.e.* $\{out(V(G' \cup Q), a) \mid G' \cup P' \in$ *preffamily*$(G, P, V, \leq, a), G' \subseteq G$, $Q \subseteq P' \subseteq P, Q$ *is a singleton or empty*$\}$.
- $out_\cup/_\cap(G, P, V, \leq)$ *are analogous as in Definition 3.*

The following example illustrates permissions as exceptions, but also obligations as exceptions to permissions.

Example 4 $G = \{a, b\}, P = \{c\}, V(a) = (\top, \neg k), V(b) = (s \wedge p, \neg k), V(c) = (s, k)$, $\{a, c\} < \{a, b\} < \{b, c\}$,[2]
It is forbidden to kill, but it is permitted to kill in case of self-defence, unless a policeman is killed.
maxfamily$(G, P, V, s) = \{\{a, b\}, \{b, c\}\}$,
preffamily$(G, P, V, \leq, s) = \{\{b, c\}\}$,
outfamily$(G, P, V, \leq, s) = \{Cn(\{\top\})\}$,
statpermfamily$(G, P, V, \leq, s) = \{Cn(\{k\})\}$

The maxfamily includes the sets of applicable compatible generators together with all non applicable ones: e.g., the output of $\{a, c\}$ *in the context s is not consistent. Even if b could conflict with c, it is not applicable in a situation s (while c is), hence* $\{b, c\}$ *is consistent; finally* $\{a\}$ *is not in maxfamily since it is not maximal, in fact, we can add the non applicable rule b. Then preffamily is the preferred set* $\{b, c\}$ *according to the ordering on set of rules above. outfamily is composed by the consequences of applying the generators whose pointer is in G which are included in* $\{b, c\}$ *which are applicable in s: b is the only obligation, but it is not applicable. statpermfamily is the result of the application of obligations in G together with a permission at a time from P, in this case c.*

[2]By $A < B$ we mean as usual $A \leq B$ and $B \not\leq A$.

5 Argumentation theory

Pollock [1987] distinguishes between rebutting and undercutting defeaters: "R is a rebutting defeater for P as a prima facie reason for Q if and only if R is a defeater and R is a reason for believing not Q", and "R is an undercutting defeater for P as a prima facie reason for [agent] S to believe Q if and only if R is a defeater and R is a reason for denying that P wouldn't be true unless Q were true." The difference between the two kinds of arguments is that "undercutting defeaters attack the connection between the reason and the conclusion rather than attacking the conclusion itself".

A well known issue within argumentation theory is the level of abstraction of the theory. We define an argumentation theory and an abstract argumentation theory. For the basic intuition of the first one, we consider arguments that consist of a single rule only. If there is a conflict between two obligations and the priority of the second obligation is higher than the priority of the first obligation, then the second obligation is a rebutter of the first one. For example the it is forbidden to have guns, but the obligation that detectives have guns has priority over this prohibition.

If there is a conflict between an obligation and a permission and the priority of the permission is higher than the priority of the obligation, then the obligation is undercut. For example, it is forbidden to kill, but the permission to kill in self defence has the priority

This intuition can be generalized and formalized as follows. An argumentation system is an input/output system with priorities. An argument is a set of norms. The priority relation on the powerset of norms becomes a priority relation on arguments.

Definition 5 *An argumentation system is a tuple $\langle G, P, V, \leq \rangle$ that consists of two disjoint sets of generator pointers G and P, a function V associating generators with pointers, and a partial pre-order $\leq$ on the powerset of $G \cup P$. A context is a propositional formula. An argument for x in context c is a minimal subset $G' \cup P'$ of $G \cup P$ such that $(c, x) \in out_{\cup}(G', P', V, \leq)$ and the set $\{x \mid (c, x) \in out_{\cup}(G', P', V, \leq)\}$ is consistent.*

A defeater of an argument $G' \cup P' \subseteq G \cup P$ for x in context c is a minimal subset $G'' \cup P''$ of $G \cup P$ such that $(c, x) \notin out_{\cup}(G' \cup G'', P' \cup P'', V, \leq)$. It is a rebutting defeater when $(c, \neg x) \in out_{\cup}(G' \cup G'', P' \cup P'', V, \leq)$, and an undercutting argument otherwise.

Example 5 *The following argument involves two parties discussing whether the defendant, a private detective, has violated some obligations since he killed a policeman by shooting him. At each turn, they put forward an obligation or a permission to prove or disprove that the defendant is guilty. These obligations and permissions compose the arguments of the parties A and B:*

1. *A: The defendant is guilty since it is forbidden to kill.*
2. *B: He did so since he was menaced, so he was permitted to kill to defend himself.*
3. *A: But he killed a policeman and it is not prohibited to kill policemen in self defence.*
4. *B: The policeman was acting against his duties, so the defendant is not guilty.*

The dispute eventually continues in this way:

1. *A: Ok, but he is guilty anyway since it is forbidden to have a gun.*
2. *B: He is a detective, so he has to bring a gun with himself.*

$G = \{a, b, e, f\}$, $V(a) = (\top, \neg k), V(b) = (s \wedge p, \neg k), V(e) = (\top, \neg r), V(f) = (t, r)$: it is forbidden to kill (k), it is forbidden to kill policemen (p) even for self defence, it is forbidden to carry firearms, but detectives (t) should bring firearms.
$P=\{c, d\}$, $V(c) = (s, k), V(d) = (s \wedge p \wedge v, k)$, in case of self defence (s) it is permitted to kill (k), in case of self defence (s) it is permitted also to kill policemen (p) who are violating their duties (v).

The partial order on the powerset of rules is the following: $\{a\} < \{c\} < \{a, b\} < \{c, d\}, \{e\} < \{f\}$

Arguments for determining that $\neg k$ is obligatory in context $p \wedge r \wedge v \wedge t$ are $G'_1 = \{a\} \cup P'_1 = \emptyset$ and $G'_2 = \{a\} \cup P'_2 = \emptyset$, because $\neg k$ belongs to the consistent sets $\{x \mid (p \wedge r \wedge v \wedge t, x) \in out_{\cup}(G'_1, P'_1, V, \leq)\}$ and $\{x \mid (p \wedge r \wedge v \wedge t, x) \in out_{\cup}(G'_2, P'_2, V, \leq)\}$.

$G''_1 = \emptyset \cup P'' = \{c\}$ is an undercutting defeater of argument $G'_1 \cup P'_1$ for $\neg k$. In fact $\neg k$ is not in the consistent set $\{x \mid (p \wedge r \wedge v \wedge t, x) \in out_{\cup}(G'_1 \cup G''_1, P'_1 \cup P''_1, V, \leq)\}$. However, $G''_1 \cup P''_1$ is not a defeater for $G'_2 \cup P'_2$, while $G''_2 = \emptyset \cup P''_2 = \{c, d\}$ undercuts it.

$G'_3 = \{e\} \cup P'_3 = \emptyset$ is an argument for $\neg r$ which is rebutted by argument $G''_3 = \{f\} \cup P''_3 = \emptyset$, because $r \in \{x \mid (p \wedge r \wedge v \wedge t, x) \in out_{\cup}(G'_3 \cup G''_3, P'_3 \cup P''_3, V, \leq)\}$.

Abstract argumentation has been popularized by Dung's work [Dung, 1995; Bondarenko *et al.*, 1997]. He defines an abstract argumentation framework as a set of arguments with a attack relation between the arguments. He thus abstracts away from context, argument for a particular formula, the structure of an argument, the distinction between rebutters and undercutters, etcetera. In our setting, both arguments and defeaters can be seen as abstract arguments, in the sense that they argue either pro and contra the possibility to infer x.

Definition 6 *An abstract argument $G' \cup P'$ is either:*

- *a subset of $G \cup P$ that is an argument for some x in some context c, or*
- *a subset of $G \cup P$ that is a defeater of some argument.*

An abstract argument α attacks another abstract argument β when α is a defeater of β.

Note that the argumentation theory adds an interesting perspective to the input/output logics framework. Whereas constraints are defined on the global level of maximal outfamilies, arguments are defined on an intermediate level between single rules and such maximal outfamilies.

6 Related work

In Pollock [1987] rebutters are a kind of defeat of an argument which attacks a reason by supporting an opposite conclusion; undercutters are arguments that attack the fact that another reason supports its conclusion: given a argument $P \rightarrow Q$, an undercutter is a reason for $\neg(P \rightarrow Q)$. However, when we consider conditionals concerning norms, as input/output logic, we should not speak of their truth. So Pollock [1987]'s definition is not adequate for obligations and permissions.

More recently, Prakken [1997] proposes a treatment of undercutters in terms of premises which contains weak negation ($\sim$) in conditionals. The argument $P \wedge \sim R \Rightarrow Q$ can be undercut by an argument supporting Q.

Verheij [2001] criticizes this approach since he argues that it limits the possibility to add further undercutting arguments once the conditional has been created. Instead, law making has a sequential character and new permissions can be added to the legal system after the obligation they are exceptions to. [Prakken, 1997] is aware of this problem and tries to overcome it by using a naming system for rules: $d : P \Rightarrow Q$. The name d is used in order to undercut a rule in a general way, by adding to the rule a premise of the form $\sim \neg Valid(d)$: if $\neg Valid(d)$ is supported then the rule d becomes not applicable anymore.

Hage [1997] proposes a different system called Reason-based logic which allows to express undercutters without introducing new premises in a rule. The operator $Excluded(thief(x) \Rightarrow punishable(x))$ applied to a rule expresses that the rule cannot be used anymore in a derivation even if it is considered valid: $Valid(thief(x) \Rightarrow punishable(x))$.

Verheij [2001] criticizes also this approach due to its complexity and proposes a different treatment of undercutters; he introduces the unary $\times\phi$ operator with the meaning that ϕ is defeated. $\phi \rightarrow \times\psi$ means that ϕ supports $\times\psi$, i.e., ϕ attacks ψ. Undercutters are represented as conditionals which defeat other conditionals; $\phi \rightarrow \times(\psi \rightarrow \chi)$ means that *phi* is an undercutter of ψ supporting χ. The interpretation of a theory in Verheij's DefLog system evaluates sentences by assigning them the values justified or defeated. A pair of sets of sentences (J, D) dialectically interprets the theory Δ when no sentence in J is both supported and attacked by other sentences in J and J attacks all sentences in D. J is called the set of justified statements and D the set of defeated statements.

In [Boella and van der Torre, 2003c] we address the notion of permission in the context of sanction based obligations. Following Boella and Lesmo [2002], in [Boella and van der Torre, 2003a] we model obligations as goals of the normative system which decides whether to sanction violations; in this context, permissions are defined as situations which are not considered violations and thus not punished [Boella and van der Torre, 2003c], following the way permissions are defined, e.g., in the Italian penal code.

7 Concluding remarks

In this paper we have discussed permissions as exceptions, and related them to rebutters in argumentation theory. The relation between permissions and undercutters raises several questions. For example, is there a distinction between weak and strong rebutters? More interestingly, we have formalized the notion of hierarchy of authorities and of *meta-norms* for resolving conflicts among authorities – see Bulygin philosophical discussion in Section 2 – in this setting, which gave rise to new notions of static and dynamic norms [Boella and van der Torre, 2003d]. This raises the question about the role of hierarchies in argumentation theory, and whether there is a distinction between static and dynamic rebutters.

In this paper we only considered the relation between permissions and undercutters, but, for example, also the following questions may be raised.

- Can permissions be related to disbeliefs, when obligations are related to beliefs?
- Can permissions be related to contractions, when obligations are related to revision?

Issues for further research come from the observation that formal argumentation has mainly focussed on deliberation about information and knowledge. Applications as finding investigations, negotiation, legal procedure and online dispute mediation, however, typically involve formal argumentation in the context of cognitive BDI and BOID agents, i.e., agents whose deliberation is based on beliefs, obligations, intentions and desires. This kind of argumentation has been pioneered by [Parsons *et al.*, 1998], [McBurney and Parsons, 2002], and in [Boella and van der Torre, 2003b] we develop the argumentation framework presented here within the BOID architecture proposed by [Broersen *et al.*, 2002].

References

[Alchourron and Makinson, 1981] C. E. Alchourron and D. Makinson. Hierarchies of regulations and their logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 125–148. Reidel, 1981.

[Alchourron, 1993] C. Alchourron. Philosophical foundations of deontic logic and the logic of defeasible conditionals. In J.-J. Meyer and R. Wieringa, editors, *Deontic Logic in Computer Science: Normative System Specification*, pages 43–84. John Wiley & Sons, 1993.

[Bobbio, 1980] N. Bobbio. Norma. In *Enciclopedia Einaudi*, volume 9, pages 876–907, Torino, 1980. Einaudi.

[Boella and Lesmo, 2002] G. Boella and L. Lesmo. A game theoretic approach to norms. *Cognitive Science Quarterly*, 2(3-4):492–512, 2002.

[Boella and van der Torre, 2003a] G. Boella and L. van der Torre. Attributing mental attitudes to normative systems. In *Procs. of AAMAS'03*, Melbourne, 2003. ACM Press.

[Boella and van der Torre, 2003b] G. Boella and L. van der Torre. BDI and BOID argumentation. In *Procs. of IJCAI Workshop on Computational Models of Natural Argument*, Acapulco, 2003.

[Boella and van der Torre, 2003c] G. Boella and L. van der Torre. Obligations and permissions as mental entities. In *Procs. of IJCAI Workshop on Cognitive Modeling of Agents and Multi-Agent Interactions*, Acapulco, 2003.

[Boella and van der Torre, 2003d] G. Boella and L. van der Torre. Permissions and obligations in hierarchical normative systems. In *Procs. of ICAIL 03*, Edimburgh, 2003.

[Bondarenko *et al.*, 1997] A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[Boutilier, 1994a] C. Boutilier. Conditional logics of normality: a modal approach. *Artificial Intelligence*, 68:87–154, 1994.

[Boutilier, 1994b] C. Boutilier. Toward a logic for qualitative decision theory. In *Procs. of KR-94*, pages 75–86, Bonn, 1994.

[Boutilier, 1994c] C. Boutilier. Unifying default reasoning and belief revision in a modal framework. *Artificial Intelligence*, 68:33–85, 1994.

[Broersen *et al.*, 2002] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly*, 2(3-4):428–447, 2002.

[Bulygin, 1986] E. Bulygin. Permissive norms and normative systems. In A. Martino and F. Socci Natali, editors, *Automated Analysis of Legal Texts*, pages 211–218. Publishing Company, Amsterdam, 1986.

[Dung, 1995] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n–person games. *Artificial Intelligence*, 77:321–357, 1995.

[Guastini, 1998] R. Guastini. *Teoria e dogmatica delle fonti.* Giuffrè, Milano, 1998.

[Hage, 1997] J.C. Hage. *Reasoning with Rules. An Essay on Legal Reasoning and its Underlying Logic.* Kluwer, 1997.

[Hansson, 1969] B. Hansson. An analysis of some deontic logics. *Nôus*, 3:373–398, 1969.

[Lewis, 1974] D. Lewis. Semantic analysis for dyadic deontic logic. In *Logical Theory and Semantical Analysis*, pages 1–14. D. Reidel Publishing Company, Dordrecht, Holland, 1974.

[Lewis, 1979] D. Lewis. A problem about permission. In E. Saarinen, editor, *Essays in Honour of Jaakko Hintikka*, pages 163–175. D. Reidel, Dordrecht, 1979.

[Makinson and van der Torre, 2000] D. Makinson and L. van der Torre. Input-output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.

[Makinson and van der Torre, 2001] D. Makinson and L. van der Torre. Constraints for input-output logics. *Journal of Philosophical Logic*, 30(2):155–185, 2001.

[Makinson and van der Torre, To appear] D. Makinson and L. van der Torre. Permissions from an input-output perspective. *Journal of Philosophical Logic*, To appear.

[Makinson, 1993] D. Makinson. Five faces of minimality. *Studia Logica*, 52:339–379, 1993.

[McBurney and Parsons, 2002] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11(3):315–334, 2002.

[Parsons *et al.*, 1998] S. Parsons, C. Sierra, and N. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.

[Pollock, 1987] J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[Prakken, 1997] H. Prakken. *Logical tools for modelling legal argument.* Kluwer Academic Publisher, Dordrecht, 1997.

[Ross, 1968] A. Ross. *Directives and norms.* Routledge and Kegan Paul., London, 1968.

[Shoham, 1988] Y. Shoham. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence.* The MIT Press, 1988.

[Verheij, 2001] B. Verheij. Legal decision making as dialectical theory construction with argumentations schemes. In *Procs. of ICAIL 01*, S. Louis (MI), 2001.

[von Wright, 1951] G. H. von Wright. Deontic logic. *Mind*, 60:1–15, 1951.

[von Wright, 1959] G. H. von Wright. *On the logic of negation.* Societas Scientiarum Fennica, Helsinki, 1959.

[von Wright, 1963] G. H. von Wright. *Norm and Action. A logical Inquiry.* Routledge and Kegan Paul, London, 1963.