OBLIGATIONS, PERMISSIONS AND TRANSGRESSIONS: AN ALTERNATIVE APPROACH TO DEONTIC REASONING

ANONYMOUS SUBMISSION
FOR THE TENTH LOGIC AND LANGUAGE SYMPOSIUM

Abstract. We present a logic of transgressions that allows deontic conflicts (i.e. conflicting obligations and permissions, and contrary-to-duty obligations (Chisholm, 1963)) but without making appeal to defeasible or paraconsistent reasoning. This logic of transgressions can be viewed as conceptually related to input/output logic (Makinson & van der Torre, 2003), except where the outputs correspond to transgressions rather than obligations. There is also a connection with the Andersonian/Kangorian reduction (Anderson, 1958; Kanger, 1971), but where the notion of a transgression is more fine-grained than a single "sanction".

In the formal analysis of obligations and permissions, one issue that arises is how to deal with "contrary to duty" obligations; those secondary obligations that we may wish to impose when primary obligations remain unfilled (Chisholm, 1963). When modelled in terms of possible worlds — where what is obligatory is what holds in some ideal worlds — this appears to require various moves to be made concerning sub-optimal worlds (Jones & Pörn, 1985, for example). Alternative approaches may require the adoption of non-standard logical notions, such as paraconsistency (Costa & Carnielli, 1986) and non-monotonicity (Bonevac, 1998, for example).

Here we propose a different approach which allows us to reason in the face of unmet obligations without resorting to notions of sub-optimality, paraconsistency, or non-monotonicity in the semantic theory. In some sense it can be thought of as a variant of input/output logic (Makinson & van der Torre, 2003), except that the output is not an indication of what is obligatory given some state of affairs, but is instead an indication of whether the input has transgressed the system of rules associated with some authority. An obedient subject may view their goal as being to minimise the number of transgressions that they make, with respect to the relevant authorities. This gives us scope to reason with obligations *even if a code of conduct is inconsistent with itself, or inconsistent with other authorities*. We also show how permissions can be formulated in such a framework.

Constraints of space mean that the following abstract can only highlight key aspects of the logic of transgressions, rather than presenting detailed illustrations of how this logic can be used to address a range of specific, problematic examples. The focus of this presentation is thus on the framework, relevant definitions (or behaviours) and a notion of deontic entailment which captures ideal patterns of inference for rational and coherent authorities.

1. Basic Notions

Assume we have some (set of, system of) obligations $\mathcal{O}_{\mathfrak{A}}$ imposed by authority $\mathfrak{A}$. If an obligation for a is part of $\mathcal{O}_{\mathfrak{A}}$, then we write $\mathsf{O}_{\mathfrak{A}}a$ (or just $\mathsf{O}a$ if $\mathfrak{A}$ is implicit). In the event that an obligation of $\mathfrak{A}$ for a is not satisfied, then a transgression has occurred, which we write $\mathcal{T}_{\mathfrak{A}}(\neg\ a)$ (dropping the subscript if if $\mathfrak{A}$ is implicit). We use $\Gamma \vdash_{\mathcal{O}_{\mathfrak{A}}} \mathcal{T}_{\mathfrak{A}}(x)$ to denote that a transgression, characterised by x, of the obligations of $\mathfrak{A}$ follows in context (state of affairs) Γ. The unfulfillability of a system of obligations is expressed by $\vdash_{\mathcal{O}_{\mathfrak{A}}} \bot_{\mathfrak{A}}$. We assume that we are dealing with *descriptions* of obligations, rather than statements that in and of themselves *impose* obligations.

Date: February 2009.

Key words and phrases. Deontic Reasoning, Authoritarian Analysis, Logic of Transgressions, Input/Output Logic.

2. Modelling Obligations

Using these notions, we can have the following definitions.

Transgression: If $\mathsf{O}_{\mathfrak{A}}a$ and $\Gamma \vdash \neg\ a$, then $\Gamma \vdash_{\mathcal{O}_{\mathfrak{A}}} \mathcal{T}_{\mathfrak{A}}(a)$. This is highly intensional: any reasoning that takes place is confined to the process of determing whether a transgression has taken place. There need be no directly specified reasoning involving the arguments of obligations as such.

Fulfillability/Coherence: If for all Γ, $\Gamma \vdash_{\mathcal{O}_{\mathfrak{A}}} \mathcal{T}_{\mathfrak{A}}(b)$ for some b, then the obligations $\mathcal{O}_{\mathfrak{A}}$ of $\mathfrak{A}$ are *unfulfillable* (and $\mathfrak{A}$ can be described as imposing *inconsistent* or *incoherent* obligations), for which we write $\vdash_{\mathcal{O}_{\mathfrak{A}}} \bot_{\mathfrak{A}}$. Otherwise they are *fulfillable* (and $\mathfrak{A}$'s obligations can be described as *consistent* or *coherent*).[1]

Vacuity: If for $\Gamma \nvdash \bot$, $\Gamma \nvdash_{\mathcal{O}_{\mathfrak{A}}} \mathcal{T}_{\mathfrak{A}}(b)$, for all b, then the obligations $\mathcal{O}_{\mathfrak{A}}$ are *vacuous* (i.e. the content of the obligations are tautologies).

3. Modelling Permissions

Now let us consider permissions. The permissions $\mathcal{P}_{\mathfrak{A}}$ of an authority $\mathfrak{A}$ could be thought of as characterising things a that should not give rise to transgressions according to $\mathfrak{A}$. If a is permitted by $\mathfrak{A}$, we can write $\mathsf{P}_{\mathfrak{A}}(a)$ (or just $\mathsf{P}(a)$ if the authority $\mathfrak{A}$ is implicit). Again we assume a descriptive rather than performative permissions, as with obligations.

We shall use (i) $\vdash_{\mathcal{P}_{\mathfrak{A}}} \bot_{\mathfrak{A}}$, (ii) $\vdash_{\mathcal{O}_{\mathfrak{A}}} \bot_{\mathfrak{A}}$, and (iii) $\vdash_{\mathcal{P}_{\mathfrak{A}},\mathcal{O}_{\mathfrak{A}}} \bot_{\mathfrak{A}}$ to denote an inconsistency in the system of (i) permissions, (ii) obligations, or (iii) both, respectively, of authority $\mathfrak{A}$.

We have the following definitions for systems of permissions $\mathcal{P}_{\mathfrak{A}}$ and obligations $\mathcal{O}_{\mathfrak{A}}$.

Consistent/Inconsistent: If $\mathsf{P}_{\mathfrak{A}}a$ and $\mathsf{P}_{\mathfrak{A}}b$, but $a \vdash \neg\ b$, then $\mathcal{P}_{\mathfrak{A}}$ is *inconsistent*, for which we can write $\vdash_{\mathcal{P}_{\mathfrak{A}}} \bot_{\mathcal{P}_{\mathfrak{A}}}$. Otherwise it is *consistent*, for which we write $\nvdash_{\mathcal{P}_{\mathfrak{A}}} \bot_{\mathcal{P}_{\mathfrak{A}}}$.

Contingently Satisfiable: If (a) $\Gamma \vdash_{\mathcal{O}_{\mathfrak{A}}} \mathcal{T}_{\mathfrak{A}}(a)$, for some Γ and a where $\Gamma \nvdash \bot$ (i.e. Γ is consistent) and $\mathsf{P}_{\mathfrak{A}}a$, and (b) $\Gamma' \nvdash_{\mathcal{O}_{\mathfrak{A}}} \mathcal{T}_{\mathfrak{A}}(b)$ for some Γ' and b where $\Gamma' \nvdash \bot$ (i.e. Γ' is consistent) and $\mathsf{P}_{\mathfrak{A}}b$, then $\mathcal{O}$ and $\mathcal{P}$ are mutually *contingently satisfiable*.

Mutually Consistent/Inconsistent: If $\Gamma \nvdash_{\mathcal{O}_{\mathfrak{A}}} \mathcal{T}_{\mathfrak{A}}(a)$ for all a such that $\mathsf{P}_{\mathfrak{A}}a$ and all Γ such that $\Gamma \nvdash \bot$ (i.e. Γ is consistent), then $\mathcal{O}$ and $\mathcal{P}$ are mutually *consistent*, for which we write $\nvdash_{\mathcal{O},\mathcal{P}\mathfrak{A}} \bot_{\mathfrak{A}}$. Otherwise they are *inconsistent*, for which we write $\vdash_{\mathcal{O},\mathcal{P}\mathfrak{A}} \bot_{\mathfrak{A}}$.

4. Conditionals

In the literature on deontic reasoning, much is made of the difficulty of analysing conditional obligations (McNamara, 2006), which has motivated the use of dyadic deontic modalities (van Fraassen, 1972, for example). Although more can be said, we will merely note that conditional obligations can be expressed in the form $p \rightarrow \mathsf{O}_{\mathfrak{A}}a$, following Castañeda (1981). We might prefer it if the notation indicated whether the conditionality was part of the authority $\mathfrak{A}$'s code. Notationally, this may be better captured by $\mathsf{O}_{\mathfrak{A}}(p \rightarrow a)$. We argue that this should still be analysed in terms of propositional detachment (the antecedent being p), rather than deontic detachment (the antecedent being $\mathsf{O}_{\mathfrak{A}}p$).

5. Complex Obligations

Related to the issue of deontic v. propositional detachment with deontic conditionals is the question of complex obligations such as *"you are obliged to help a robbed man"* (Prior, 1958), where we do not (in this case) wish the obligation distribute down to the property of being robbed (that is, there should not be an obligation to rob). We might question whether the intuitions behind such examples are strong enough for them to be hard-coded in any semantic theory, as opposed to being contingent, pragmatic artifacts of a pre-existing moral presumption, namely, that it is wrong to rob. We only need to consider other less emotive obligations,

[1] Hansen *et al.* (2007)Note discuss the notion of coherence, among other things, and how it is best defined.

such as *"you are obliged to use a clean knife"* for us to begin to question the fundamental nature of such intuitions.[2]

To some extent the issues are independent of the particular features of the logic proposed here; arguably, this is more a question of the appropriate syntax-semantics interface, perhaps combined with an element of pragmatic reasoning.

6. Entailments

We may now wish to consider the nature of entailments that are or should be supported by a deontic system expressed in terms of a logic of transgressions. In the current formulation, this includes entailments and associated behaviours between expressions of the following forms:

(1) $\mathsf{O}a$ and $\mathsf{O}b$;
(2) $\mathsf{P}a$ and $\mathsf{P}b$;
(3) $\mathcal{T}(a)$ and $\mathcal{T}(b)$;
(4) $\mathsf{O}a$ and $\mathsf{P}b$;

where there is some inferentional connection between a and b according to their interpretation in classical or other logics.

If we assume a classical logic, we can consider the nature of any inferential behaviour governing obligations and permissions by using *reductio* arguments. For example, to determine whether $\mathsf{O}(a)$ "follows from" $\mathsf{O}(a \wedge b)$ we can consider whether adding $\neg\ (\mathsf{O}(a \wedge b) \rightarrow \mathsf{O}(a))$ leads to unfulfillability.[3]

More generally, we have a notion of "inference" (or "presumption") $A \Longrightarrow B$ meaning that if we have A and it is fulfillable/consistent (or contingently acceptable in a given context Γ), adding $\neg\ B$ leads to unfulfillability/inconsistency (or a transgression in Γ). It is important to realise that a system of obligations and permissions need not be consistent or fulfillable, and these kinds of inferences do not have to be endorsed by any authority. In effect, these kinds of inferences indicate what should follow if we assume that an authority is rational and has consistent and fulfillable demands.

7. Obligation, Permission and Negation

One issue of particular interest when comparing this system with Standard Deontic Logic (SDL) (see McNamara (2006) for example) is the nature of the relationship between permissions and obligations — for example by way of transgressions, fulfillability and consistency — and whether and in what way they are interdefinable. For example, we may wonder if, and in what way, this logic of transgressions supports the following.

$$\mathsf{O}a \rightarrow \mathsf{P}a$$

It is also natural to consider how to interpret negated deontic expressions in this setting, giving the usual definitions for the interdefinability of obligation and permission in SDL and related formalism, i.e. where $\mathsf{P}a =_{def} \neg\ \mathsf{O}\ \neg\ a$, and the interpretation of statements such as *"You are not permitted to close the door"*, *"You are not obliged to open the window."*

This latter definition requires us to consider the intended meaning of negation in the context of obligations and permissions. There are a range of answers to such questions. Here there is only sufficient space to note that the following may be taken to correspond to the above axiom and definition, respectively.

$$\mathsf{O}a \Longrightarrow \mathsf{P}a$$

$$\mathsf{P}a \Longrightarrow \neg\ \mathsf{O}\ \neg\ a \text{ and } \neg\ \mathsf{O}\ \neg\ a \Longrightarrow \mathsf{P}a$$

Of course, to justify the claim that such statements capture the desired intuitions, we need to know how negated obligations and permissions are to be interpreted, and whether such negation corresponds to a natural interpretation of negated obligations and permissions in

[2]This rather subtle topic is the subject of another paper by the current author.

[3]There is a sense here in which something does not "hold" if it leads to a transgression that is not so unlike the constructive notion of negation, where a proposition p is false if we can derive $\perp$ from an assumption that p is true ($\neg\ p$ iff $p \vdash \perp$).

natural language. For example, for an authority to say that *"you are not obliged to p"* may be different to the mere absence of the obligation to p (similarly for permission). Depending on how this is formulated, it is possible to produce a system that supports normative gaps, addressing a long-standing puzzle for Standard Deontic Logic (von Wright, 1968),

8. Conclusions and Future Work

The paper presents a novel approach for capturing aspects of the semantics of obligations and permissions in a way that allows us to deal with unmet obligations and inconsistent codes of conduct without resorting to the machinery of paraconsistent logic or non-monotonic reasoning.

The treatment provides a rich, fine-grained framework in which issues concerning negation and entailment, and various deontic puzzles can be considered without sometimes inappropriate solutions being imposed or required by a conventional possible-worlds style account.

References

Anderson, Alan Ross (1958), A reduction of deontic logic to alethic modal logic, *Mind* 67:100–103.

Bonevac, Daniel (1998), Against conditional obligation, *Noûs* 32(1):37–53.

Castañeda, Hector Neri (1981), The paradoxes of deontic logic: The simplest solution to all of them in one fell swoop, Reidel, Dordrecht, ISBN 9027713464, 9789027713469, (37–85).

Chisholm, Roderick M. (1963), Contrary-to-duty imperatives and deontic logic, *Analysis* 24:33–36.

Costa, Newton C. A. Da & Walter A. Carnielli (1986), On paraconsistent deontic logic, *Philosophia* 16(3–4):293–305.

van Fraassen, Bas (1972), The logic of conditional obligation, *Journal of Philosophical Logic* 1(3/4):417–438, iSSN: 0022-3611 LC: BC51 .J68.

Hansen, Jörg, Gabriella Pigozzi, & Leendert van der Torre (2007), Ten philosophical problems in deontic logic, in Guido Boella, Leon van der Torre, & Harko Verhagen (eds.), *Normative Multi-agent Systems*, Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, Dagstuhl, Germany, 07122, ISSN 1862-4405.

Jones, Andrew & Ingmar Pörn (1985), Ideality, sub-ideality and deontic logic, *Synthese* 65:275–290.

Kanger, Stig (1971), New foundations for ethical theory, in Risto Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings*, D. Reidel., Dordrecht, (36–58), originally published as "New Foundations for Ethical Theory, Part I" in 1957, Stokholm.

Makinson, David C. & Leendert van der Torre (2003), What is input/output logic?, in *Foundations of the Formal Sciences II: Applications of Mathematical Logic in Philosophy and Linguistics*, Kluwer Academic Publishers, Dordrecht, volume 17 of *Trends in Logic series*, (163–174).

McNamara, Paul (2006), Deontic logic, in Dov Gabbay & John Woods (eds.), *The Handbook of the History of Logic*, Elsevier Press, Amsterdam, volume 7, Logic and the Modalities in the Twentieth Century.

Prior, A. N. (1958), Escapism: The logical basis of ethics, in A.I. Melden (ed.), *Essays In Moral Philosophy*, University of Washington Press, Seattle, (135–146).

von Wright, G. H. (1968), *An Essay in Deontic Logic and the General Theory of Action*, North Holland Publishing Co.