Why be Afraid of Identity?
Comments on Sergot and Prakken's views

Xavier Parent

Individual and Collective Reasoning (ICR) Group
Computer Science and Communications (CSC)
Faculty of Sciences, Technology and Communication (FSTC)
University of Luxembourg
6, rue Richard Coudenhove - Kalergi
L-1359 Luxembourg
`xavier.parent@uni.lu`

Abstract. The paper discusses the views held by Sergot and Prakken [22] on the import, or non-import, of the identity principle for conditional obligation within a preference-based semantics. This is the principle $\bigcirc(A/A)$. The key point is to understand and appreciate what unconditional obligations the principle allows us to detach, and from what premises. It is argued that it does not license the move from A to $\bigcirc A$, which would amount to committing a breach of Hume's law: no 'ought' from 'is'. It is also shown that the most that is licensed is the move from $\Box A$ to $\bigcirc A$ – a move that appears to be harmless, and (above all) compatible with the idea that obligations are essentially violable entities. An existing pragmatic theory can be used to explain it. Objections based on the definition of the unconditional obligation operator are countered.

Keywords: Conditional obligation; preferences; identity; detachment; Hume's law

1 Introduction

Three principles were singled out by tradition under the name "laws of thought". These are:

- The law of identity: "If a proposition is true, then it is true."
- The law of non-contradiction: "A proposition cannot be both true and not true at the same time."
- The law of excluded middle: "Every proposition is either true or false."

These three laws are no longer singled out in quite this way. In recent years they have been under assault and dismissed by a number of logicians. This is particularly true of the laws of non-contradiction and excluded middle. These are denied by paraconsistent logic and intuitionistic logic, respectively. It is sometimes believed that only the law of identity has remained unchallenged. As a matter of facts, it has also come under criticism for not being suitable for dealing with causal relationship (see, e.g., Shoham [24, p. 218]).

This paper focuses on another area where the desirability of the principle of identity has been discussed. It is the area of deontic logic, which deals with obligations and permissions. The fact that identity has been an issue there is less known. Expressed in terms of conditional obligation, the identity principle is the law $\bigcirc(A/A)$. This is usually read as "If A is the case, then it is obligatory that A". It would look strange to have it in the logic. Some well-known systems do have it, and there are deontic logicians who have been worried by this. A quite interesting defense of the view that the worry is misconceived is offered by Sergot and his co-author Prakken in [22]. In this paper I wish to explain how I understand their suggestion, and discuss their view. The point they make may seem a rather small one, but I think it is a valuable one, and I have not seen it discussed in any depth in the deontic logic literature.

This paper is organized as follows. Section 2 provides the necessary background for the discussion. Section 3 explains Sergot and Prakken's argument. Section 4 discusses two series of objections that may be raised in relation to it. Section 5 concludes.

2 Background

First, some background is necessary. I shall start by explaining how the above problem arose for Sergot & Prakken. Lurking in the background there is the Hansson/Lewis preference-based semantics for dyadic deontic logic ([13, 16]). It emerged in the early 70s as one of the most suitable tools for dealing with norm violation and contrary-to-duties (CTDs),[1] and nowadays it seems to remain very popular, especially amongst philosophers. The key idea is to interpret "It ought to be that B given A" – $\bigcirc(B/A)$– as true if and only if the best A-worlds are all B-worlds. In their [22], Sergot and Prakken convincingly show that the Hansson/Lewis semantics must be extended in order to give a more fine-grained analysis of CTDs structures. I will not go into the details of their proposal. For present purposes, I just wish to point out that their extension is a conservative one, in the sense that most, if not all, of the formulas that are validated under the the original account remain so under the extended one. Amongst them is the identity principle. It holds, because the best A-worlds are all A-worlds. Its counting as a law under the Hansson/Lewis account was reckoned very early. This has often been used even as an objection against the account (see, e.g. [2, p.95]). To make their approach to CTDs viable, Sergot and Prakken need to counter the objection.

There is a more general issue at stake here. It is the question of whether deontic logic is a viable tool for modeling normative multi-agent systems. I can-

[1] A contrary-to-duty obligation says what should be done if another (primary) obligation is violated. Hereinafter I will on some occasions use the term "according-to-duty" (ATD, for short) to refer a conditional obligation that tells us what should be done if another obligation is fulfilled.

not more agree with Sergot when he makes the point that "agent interactions generally can best be regulated and described by the use of norms" [23, p.1]. Norms provide a means for increasing coordination amongst agents. Furthermore, it would be a mistake to assume that agents always behave the way they should. Thus, the possibility of norm violation must be kept open. By clearing the charge leveled against the Hansson/Lewis account, they provide further evidence that deontic logic has a potentially useful role to play in the study of normative multi-agent systems.

To make this paper self-contained, I briefly recall the semantics of so-called dyadic deontic logic. Formulas of the form $\bigcirc(B/A)$ are interpreted in models of the form $M = (W, \succeq, \iota)$ where

- W is a set of possible worlds $w, w', ...$
- $\succeq \subseteq W \times W$ is a pre-order, i.e. it is a reflexive and transitive relation on W; $w \succeq w'$ can be read as "w is at least as good as w'"
- ι is a function which assigns to each propositional letter p in Prop a subset of W; intuitively $\iota(p)$ is the set of worlds where p is true

Let $[A]_M$ denote the truth-set of A in M, i.e. the set of worlds in M at which A is true. I will omit the subscript M when it is clear which model is intended. Let $\max_{\succeq}([A])$ denote the set

$$\{w \in [A]_M \mid \forall w'(w' \in [A]_M \,\&\, w' \succeq w) \rightarrow\ w \succeq w'\}$$

Intuitively $\max_{\succeq}([A])$ is understood to be the set of best A-worlds.[2] The truth-conditions for the dyadic obligation operator read:

(1) $M, w \models \bigcirc(B/A) \Leftrightarrow \max_{\succeq}([A]) \subseteq [B]_M$

According to (1), world w in model M satisfies the obligation "B should be the case if A is the case", $M, w \models \bigcirc(B/A)$, whenever in the best worlds, where A holds, B holds too. $\bigcirc A$ is taken to be equivalent to $\bigcirc(A/\top)$ – where $\top$ is a tautology.

The validity of the identity principle follows from the definitions being used. We have $\max_{\succeq}([A]) \subseteq [A]$. Usually authors take for granted that the validity of $\bigcirc(A/A)$ is counter-intuitive, and take pain to explain how they think it can be blocked. Sergot and Prakken should be given credit for raising the problem of whether it should be blocked in the first place. This is a preliminary issue, which must be clarified first. Furthermore, there are grounds to believe that some of the proposed solutions just do not work. I am referring to the solutions based on time. For instance, Makinson [17] suggested making the futurity dimension explicit in the formalism. In my [20] I already explained why the proposal is not convincing. A related, albeit distinct, approach has been taken by both Spohn [26, p.250] and Alchourrón [2, p.95]. One might refer to it as the "time lag" idea.

[2] The set of best A-worlds is usually assumed to be non-empty whenever A is consistent. This is known as the "limit assumption". Such an assumption is not germane to my purposes, and thus I will ignore it.

The key point is to require that the condition of a conditional obligation should occur before the obligatory formula. Incidentally a somewhat similar idea has been put forth in the context of the analysis of CTDs. In [21, p. 92] and [22, p. 244] Sergot and Prakken (rightly, in my view) object that there are obligations whose consequent occurs at the same time as its antecedent. These are quite common, and thus a logic implementing the requirement would be very limited in scope.

It may be objected that conditional obligations of which antecedent and consequent refer to the same time can still be reincorporated in the logic, by formalizing them as (unconditional) obligations of a disjunction. One of Sergot and Prakken's favorite examples involves the following two norms: there must be no fence; if there is a fence, it must be white. Under this proposal, the first (or primary) obligation would be rendered as the (unconditional) obligation that there is no fence, $\bigcirc\neg f$, and the second obligation would be rendered as the (unconditional) obligation that either there is no fence or it is white, $\bigcirc(\neg f \vee w)$. A drawback of the proposal (and I think a good reason for not pursuing it further) is that the contrary-to-duty obligation follows logically from the corresponding primary obligation. If the best worlds are all $\neg f$-worlds, then *a fortiori* they are all $\neg f \vee w$-worlds.

There is an inference pattern that takes center stage in their discussion of the import of the identity principle. It is the rule of factual detachment

$$\text{(FD)}\ \frac{\bigcirc(B/A) \qquad A}{\bigcirc B}$$

The reader needs to appreciate fully the importance of this rule (and related ones) for deontic reasoning. If one does not, one might fail to get their point. Most systems of deontic logic have this law, perhaps in a qualified form. There are two good reasons for this.

First, a logic that does not allow "deconditionalization" is, I think, of little practical use. I cannot more agree with van Eck when he asks: " How can we take seriously a conditional obligation if it cannot, by way of detachment, lead to an unconditional obligation?" [7, p.263]. Obligations and permissions are contextual and vary based on the setting. Consequently, a norm always takes the form of a conditional statement. However, in the notation $\bigcirc(B/A)$, the antecedent A has the nature of a hypothesis, which needs to be discharged for the obligation to apply. Thus, a deontic logic that does not allow detachment will never lead to an action, because the agent will never detach any conclusion about what is obligatory, and act according to what he believes to be the best.

Second, some philosophers like [4] think (rightly, in our view) that the disposition to reason according to detachment is constitutive of the possession of the concept of conditional, and thus of the concept of norm. The idea is that, if someone says "if A then obligatorily B", and if he truly means it, then he commits himself to detaching the (unconditional) obligation of B given A. If he refuses to acknowledge that he is justified in employing detachment, this will be good evidence that he fails to understand what is meant by "if ... then".

Accepting detachment and acquiring an implication are simply two sides of the same coin.

3 Sergot and Prakken's argument

Now I can get back to Sergot and Prakken's argument. The reason why the validity of $\bigcirc(A/A)$ may look counter-intuitive is that the law converts any fact into an obligation. There are two things that make it look like counter-intuitive. For one, the law applies to *any* state of affairs A. For another, it exhibits a category error, since it moves from an "is" statement to an "ought" statement. This violates what is usually referred to as Hume's law, which states that you can not derive an "ought" from an "is".[3] There is widespread agreement amongst philosophers that such a move is not possible.

Still, Sergot and Prakken give us to understand that this is all right. Following one of Hansson's suggestions, they argue that

(i) the validity of $\bigcirc(A/A)$ is not paradoxical

if we assume that

(ii) the antecedent of any conditional obligation describes some "settled" (fixed) situation.

Here, the notion of settledness is captured by means of some alethic S5 modal operator $\Box$, and it must be understood in an objective sense. The evaluation rule for it is standard: $\Box A$ holds at world x if A holds in all the possible worlds. They do not say much by way of further characterization of 'objectively' necessary, except that they want to distinguish it from settledness in a subjective sense, "such as when an agent decides to regard it as settled for him that [this-or-that will be the case]" [22, p. 241].

One might justify (ii) by contrasting the factual detachment rule alluded to earlier with its variant called "strong factual detachment":

$$\text{(FD)}\ \frac{\bigcirc(B/A) \qquad A}{\bigcirc B} \qquad\qquad \text{(SFD)}\ \frac{\bigcirc(B/A) \qquad \Box A}{\bigcirc B}$$

It turns out that, of these two inference patterns, only the second holds in the Hansson/Lewis semantics. In this respect, (ii) can be taken to mean

(iii) an agent cannot detach any conclusion about what is obligatory − and act according to what he believes to be the best, unless he believes the antecedent to be settled as true.

If we assume (ii) in the sense given by (iii), then it becomes natural to explain (i) as follows. Substituting A for B in (SFD), and invoking the validity of $\bigcirc(A/A)$, we immediately get :

[3] The law is so named, because it was originally phrased by Scottish philosopher David Hume [14].

$$(\star)\ \frac{\Box A}{\bigcirc A}$$

Since (FD) fails, we do not have:

$$(\#)\ \frac{A}{\bigcirc A}$$

A logic that allows (#) as a rule of inference can be described as committing a breach of Hume's law. For (#) enables us to derive an "ought"-conclusion from any "is"-premise. $\bigcirc(A/A)$ has this import only in the case of unalterable facts. A natural reaction is to regard this as harmless: in the context of $\Box A$, $\bigcirc A$ is redundant, and has no import. That is how I understand Prakken and Sergot's line of defense, when they write:

> "$\bigcirc(A/A)$ says only that A holds in all of the best accessible A-worlds, which is no more (or less) unacceptable than the validity of $\bigcirc\top$ in standard deontic logic ... if the context is objectively settled, then the truth of A is unalterable ; again there seems nothing particularly odd about saying that what is unalterably true is also obligatory".[4] [22, p. 245]

A tautologically true state-of-affair is the prototype of an unchangeable fact; hence the analogy with the validity of $\bigcirc\top$ in standard deontic logic. Unlike them, I would not claim that there is nothing odd about saying that what is unalterably true is also obligatory. Rather, I would concede that you may have an opinion that there is some oddity in saying it. But what I think makes the oddity acceptable is that it is of pragmatic character, not of semantic character.

4 Assessing the argument

This section discusses two series of objections to Sergot & Prakken's argument.

4.1 First series of objections

The first series of objections centers around the rules (SFD) and ($\star$). In an attempt to clear the charge leveled against the identity principle, Prakken and Sergot argue that the most that comes from it is the law ($\star$). The strategy succeeds only if we accept (SFD). Non-monotonic researchers might find this inference pattern somewhat "unusual". What they are used to is the defeasible variant of modus ponens. It warrants the inference from "A" and "If A then B" to "B" in the absence of information to the contrary only, and thus the inference rule is defeasible. This is useful to handle exceptions. The conditional "if A then B" is here modified by the qualifier "normally". However, it should

[4] The first formula in the quotation has been adapted to the notational convention used here. I use the notation $\bigcirc(B/A)$ where they use $\bigcirc[A]B$.

be recalled that Prakken and Sergot are primarily concerned with the analysis of contrary-to-duty obligations, which tell us what comes into force when some other obligations are violated. It should also be kept in mind that they (rightly, in my view) take it as crucial not to confuse violation and exception. The primary obligation to pay one's income tax might well leave room for exceptions. But, when such an exception occurs, it makes no sense to say that the obligation is violated. With this in mind, one might find (SFD) easier to understand. It means the following:

> "As long as it is possible to avoid violation of a primary obligation $\bigcirc\neg A$ a CTD obligation $\bigcirc(B/A)$ remains restricted to the context [and cannot be deconditionalized]; it is only if the violation of $\bigcirc\neg A$ is unavoidable, if $\Box A$ holds, that the CTD obligation comes into full effect, pertains to the context $\top$". [22, p. 241]

Still, the question of whether (SFD) is a reasonable inference pattern is a contentious issue. Jones and Carmo make the following objection:

> "if $\Box A$ holds, then $\bigcirc A$ holds, and thus – since the (D)-schema is valid in their system – $\neg \bigcirc \neg A$ holds. That is, going back to the dog-and-sign scenario, one could detach the obligation to put up a sign only in circumstances in which there was no longer an obligation not to have a dog." [6, p.338]

The dog-and-sign scenario, to which reference is made here, involves the following three norms: there should be no dog; if there is no dog, there should be no warning sign; if there is a dog, there should be a warning sign. These are rendered as $\bigcirc\neg d$, $\bigcirc(\neg s/\neg d)$ and $\bigcirc(s/d)$, respectively. Carmo and Jones point out that, for $\bigcirc s$ to be detached from the third, not only $\Box d$ must hold, but also $\neg \bigcirc \neg d$. Hence, the price we pay for keeping the identity principle is that, once violated, a primary obligation (here, the prohibition of A, having a dog) ceases to be in force. However, the objection can be countered. For Sergot and Prakken may reply that this is precisely their point: the state of affairs A is unavoidable. There is some oddity in saying that A is no longer forbidden, but again this is harmless.

Foreseeing a reply along these lines, Jones and Carmo make a second move:

> "The notion of 'settledness' Prakken and Sergot employ is peculiar, at least with respect to its relation to the concept of obligation. How can that which is settled, unalterable, be obligatory? Surely that which is genuinely obligatory must be violable! Here again is a basic point of contrast between our approach and that of Prakken and Sergot." [6, ibidem]

Here Carmo and Jones give us to understand that ($\star$) must be rejected, because it is incompatible with the seemingly plausible idea that an obligation is violable. The argument can be countered again. To the question of whether the two are incompatible, I say "yes" and "no". I say "yes", because one cannot adopt ($\star$) along with the law (where $\Diamond$=possibly)

$$(\star\star)\ \frac{\bigcirc A}{\Diamond\neg A}$$

$(\star\star)$ says that "ought" entails "possibly not" in the sense of physical possibility. For, putting $(\star)$ and $(\star\star)$ together we arrive at the absurdity that, if A is settled as true, then it is not settled as true:

1. $\Box A$	Assumption
2. $\bigcirc A$	1, by $(\star)$
3. $\Diamond\neg A$	2, by $(\star\star)$
4. $\neg\Box A$	3, by duality

But I say "no", because one might still reply that $(\star\star)$ must be viewed as a conversational implicature (in the sense of Grice [12]) rather than a logical inference. A fully worked out theory of conversational implicatures like that of Gazdar [10] can be used to substantiate this claim further. If $(\star)$ holds, then $\Box$ and $\bigcirc$ form what Gazdar calls a "scale", $\langle\Box,\bigcirc\rangle$. Under his account, to assert the weaker element, $\bigcirc A$, conversationally implicates that (the speaker knows that) the stronger does not hold, $\neg\Box A$.[5] Thus,"ought" does not logically entail "possibly not", but instead conversationally implicates it.[6] Earlier I mentioned the semantics-pragmatics distinction to explain away the oddity of saying that what is unalterably true is also obligatory. I am fine with the idea of invoking pragmatic considerations for corrective purposes, like in the old days. I am fine with it, especially if a relatively well-accepted pragmatic theory can back you up by predicting the peculiar phenomenon you are concerned with.

4.2 Second series of objections

The second series of objections centers around the equivalence $\bigcirc A \leftrightarrow \bigcirc(A/\top)$. Sergot & Prakken's argument has it as premise. The equivalence is involved in (SFD). This is the traditional way to define the unconditional obligation operator in terms of the conditional obligation one. As Makinson [18] observes, this definition amounts to saying that A is obligatory if it is so under zero information about the world. He calls this the minimal definition of "ought". However, there are alternative ways to define the unconditional obligation operator in terms of the conditional one. It is natural to ask if Sergot & Prakken's argument still goes through if such an alternative definition is used. This section is devoted to answering this question.

Objection based on the actual world account Von Wright [27] takes $\bigcirc A$ as an abbreviation for $\bigcirc(A/\sigma)$, where σ is a propositional constant standing for the actual world. I shall refer to it as the actual world definition of "ought". Such

[5] A good exposition of Gazdar's theory can be found in Levinson [15].

[6] Sinnot-Amstrong [25] makes a similar point concerning the relationship between "ought" and "can".

a definition is implemented within a Hansson-type semantics by Alchourrón [1]. A model M is as before, but (like in Kripke's formulation of modal logic) its universe W comes supplemented with a distinguished actual world "$\star$", such that $\star$ is the only world where σ holds:

(2) $[\sigma]_M = \{\star\}$

And a formula is defined as valid in a model M if it is true at the model's actual world $\star$. On the proof-theoretical side, the characteristic rule of σ is Meredith's law [7]

$$\frac{A}{\Box(\sigma \rightarrow A)}$$

On the other hand Hansson's logic has the rule known as "weakening the consequent":

$$(\text{WC}) \frac{\bigcirc(B/A) \qquad \Box(B \rightarrow C)}{\bigcirc(C/A)}$$

It is not difficult to see that, given identity, a breach of Hume's law immediately follows:

1. A	Assumption
2. $\Box(\sigma \rightarrow A)$	1, by Meredith's law
3. $\bigcirc(A/\sigma)$	2, by $\bigcirc(\sigma/\sigma)$ and WC
4. $\bigcirc A$	3, by definition

Thus, at first sight, it would seem that Sergot & Prakken's argument does not go through if the actual world definition is used: the logic allows (#) as an inference rule, and the identity principle is clearly involved in this.

In my view, this objection does not stand up to close scrutiny. For there exist some independent reasons against the actual world proposal. It may be argued that the definition remains too strong in itself. It corresponds to what Makinson calls the "maximal" [18] notion of unconditional obligation. As opposed to the minimal notion, which puts A obligatory under zero information about the world, the maximal notion puts A as obligatory under complete information about the word. In the notation "$\bigcirc(A/\sigma)$", the antecedent σ can be viewed as the conjunction of all boolean formulas presently true. Hence the validation of Meredith's law.

This leads to another "incongruity". It is that, under the proposed account, the move from $\bigcirc A$ to A is also warranted so that the unconditional obligation operator in fact collapses. This can be verified directly, by semantic methods. Assume $\bigcirc A$ holds at the actual world $\star$. By (1) $\max_{\succeq}([\sigma]) \subseteq [A]$. By (2), the definition of max and reflexivity of $\succeq$, $\{\star\} = [\sigma] \subseteq \max_{\succeq}([\sigma])$. Putting the two together, we get the result $\{\star\} \subseteq [A]$, which suffices for A to be true at $\star$. This makes the proposal all the more suspect to me.

[7] The rule is named after Meredith, who sees it as the most salient feature of the Wittgensteinian notion of "world" in the sense of "everything that is the case" (see [19]).

Objection based on the existential account Makinson [18] mentions the possibility of defining an intermediate notion of unconditional obligation, which would put A as obligatory under current information about the world. He adds he is not aware of any formal accounts in the literature despite its evident importance. I believe one is available. A weaker definition is proposed by Åqvist [3]. His point of departure is the approach to propositional quantification in modal logic presented by Fine [9]. The Hansson/Lewis account of conditional obligation is combined with the latter approach, and $\bigcirc A$ is defined as an abbreviation for $\exists p\ (p \wedge \bigcirc(A/p))$. For ease of exposition, I will present the simplified version of the resulting framework given by Bombosch [5], who chooses not to make the propositional quantifiers part of the object-language. To help with cross reference, I shall call this the existential account of "ought". On the semantic side, a model becomes a quadruplet $M = (W, P, \succeq, \iota)$, where W, $\succeq$, and ι are as before, and P is a nonempty subset of $\mathcal{P}(W)$. P should be considered as the set of propositions or states-of-affairs. The assumption is, then, made that

(3) $[A]_M \in P$ for every formula A

The evaluation rule for the unconditional obligation operator runs:

$$M, w \models \bigcirc A \Leftrightarrow \text{there is some state of affairs } X \in P, \text{ such that } w \in X \text{ and } \max_{\succeq}(X) \subseteq [A]_M$$

It is not difficult to show that the rule of detachment is now available in its factual form, (FD). For suppose $\bigcirc(B/A)$ and A both hold at w. From the first, $\max_{\succeq}([A]_M) \subseteq [B]_M$. By (3), $[A]_M \in P$. Since $w \in [A]_M$, one might appeal to the evaluation rule for $\bigcirc$ to conclude that $\bigcirc B$ holds at w.

A new objection to Sergot & Prakken's argument can now be made. It looks as though their argument does not go through if the existential account is used. For the reason explained in Section 3, given (FD) the logic warrants the inference from A to $\bigcirc A$, and the identity principle is clearly involved in this.[8]

The objection can be countered again. I believe that there are independent reasons to be uneasy with the existential account as it stands. Basically I have a concern about how the account handles the usual CTDs scenarios, like the dog-and-sign scenario mentioned above.[9] The logical representation of the scenario is $\{\bigcirc\neg d, \bigcirc(\neg s/\neg d), \bigcirc(s/d), d\}$. Figure 1 shows a typical model of this premises set.

---

[8] The fact that, given an arbitrary formula A, $A \rightarrow \bigcirc A$ is a valid statement, is pointed out by Bombosch [5, p. 157] too.

[9] This issue is not the prime focus of Åqvist in [3]. There he is mainly concerned with giving an account of so-called *prima facie* obligations. A prima facie obligation is an obligation that is binding all other things being equal, that is, unless it is overridden by another obligation. For a review and discussion of the state of the art on this topic, the reader is referred to Goble [11].

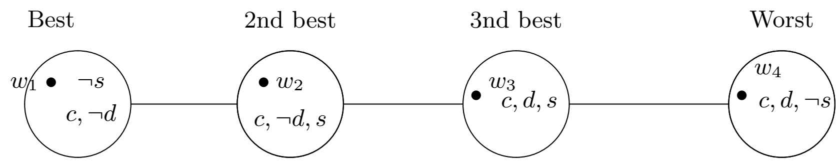

Fig. 1. Dog-and-sign scenario

Some suitably chosen propositional letter c is used. Intuitively, c denotes the relevant circumstances that make the obligations apply: I rent a given flat, the landlord does not want dogs because he fears that he may be financially responsible for damage or injury caused by the dog, etc. It is assumed that the actual world is w_3, and that $P = \mathcal{P}(W) - \emptyset$. d holds in w_3, and so does $\bigcirc\neg d$. This is because c holds at w_3, and the best c-world makes d false. $\bigcirc(\neg s/\neg d)$ and $\bigcirc(s/d)$ also hold at w_3, because the best $\neg d$-world makes s false, and the best d-world makes s true. This shows that all the sentences in the premises set are satisfied.

It is instructive to see what other obligations are made true. $\bigcirc s$ holds at w_3. This is because $c \wedge d$ holds at w_3, and the best $c \wedge d$-world makes s true. But $\bigcirc\neg s$ also holds at w_3, because e.g. c holds at w_3, and the best c-world makes s false.

In this example, the obligations $\bigcirc s$ and $\bigcirc\neg s$ both hold in w_3. Intuitively, the second obligation looks counter-intuitive. Thus, the existential account overgenerates certain unconditional obligations. It makes an ATD obligation detachable in a violation context even though it looks as though it should not be made so. The account fails in this regard, and this can be viewed as a shortcoming. For this reason, I do not believe that (in its current form) the latter account can be used to undermine Sergot & Prakken argument.

5 Conclusion

In the end, I fully endorse the views held by Sergot and Prakken on the import of the identity principle for conditional obligation, $\bigcirc(A/A)$, within a preference-based semantics. The key point is to understand and appreciate what unconditional obligations the principle allows us to detach, and from what premises.

A closer look reveals that the logic does not license the move from A to $\bigcirc A$, which would clearly be undesirable. For this would amount to committing a breach of Hume's law: no 'ought' from 'is'. Such a breach would be inescapable in the presence of additional laws, which link the conditional obligation operator to the unconditional one. These additional laws are not part of the account, and for good reasons.

An even closer look reveals that the most that is licensed is the move from $\Box A$ to $\bigcirc A$. Such a move is harmless, because the agent cannot change the truth of A. This inference pattern remains compatible with the idea that obligations are essentially violable entities. An existing pragmatic theory like Gazdar's one can be used to explain it. Under this account, to assert "it ought to be that A" conversationally implicates that the agent knows that the obligation can be violated.

6 Afterword

Marek Sergot has contributed greatly to our understanding of the key issues in deontic reasoning. He spent considerable effort on making the intricacies surrounding the problem of reasoning about norm violation (the CTD problem, for short) clearer to us. It is nearly impossible to summarize his contributions to the subject in a few words but here are some highlights. His most cited publications on CTDs are two joined papers with Henry Prakken, [21] and [22]. There they compile a set of benchmark examples − some were drawn from previous literature and some they invented themselves − against which a satisfactory account of CTDs should be assessed. They argue that the Hansson/Lewis approach based on an ordering of possible worlds, in terms of relative goodness, fares better than other approaches available from literature. They also pinpoint some shortcomings of the account, and offer a remedy for them. The basic idea is to give a special status to designated explicit obligations in the premises set. These developments are powerful, but remain tangential to the main thrust of the present paper, and have no bearing on its conclusions. This is why they have been put to one side.

Acknowledgments. I wish to thank three anonymous referees for helpful comments.

References

1. C. Alchourrón. Philosophical foundations of deontic logic and the logic of defeasible conditionals. In J.J. Meyer and R.J. Wieringa, editors, *Deontic Logic in Computer Science*, pages 43–84. Wiley and Sons, New York, 1993.
2. C. Alchourrón. Defeasible logic: Demarcation and affinities. In G. Crosso, L. Fariñas del Cerro, and A. Herzig, editors, *Conditionals and Artificial Intelligence*, pages 67–102. Clarendon Press, Oxford, 1994.
3. L. Åqvist. Prima facie obligations in deontic logic: A Chisholmian analysis based on normative preference structures. In Fehige and Wessels [8], pages 135–155.
4. P. Boghossian. Knowledge of logic. In P. Boghossian and C. Peacocke, editors, *New Essays on the A Priori*, pages 229–254. Clarendon Press, Oxford, 2000.
5. U. Bombosch. The meaning of 'ought, prima facie' and decision situations: A reply to Aqvist. In Fehige and Wessels [8], pages 152–162.
6. J. Carmo and A.J.I. Jones. Deontic logic and contrary-to-duties. In D. M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 8, pages 265–344. Kluwer Academic Publishers, Dordrecht, Holland, 2nd edition, 2002.

7. J. Van Eck. A system of temporally relative modal and deontic predicate logic, and its philosophical applications. *Logique & Analyse*, 25:249–290 and 339–381, 1982.
8. C. Fehige and U. Wessels, editors. *Preferences*. De Gruyter, Berlin - New York, 1998.
9. K. Fine. Propositional quantifiers in modal logic. *Theoria*, 36:336–346, 1970.
10. G. Gazdar. *Pragmatics: Implicature, Presupposition, and Logical Form*. Academic, New York, 1979.
11. L. Goble. Prima facie norms, normative conflicts and dilemmas. Forthcoming in [**?**].
12. P. H. Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics, Volume 3: Speech Acts*, volume 3, pages 41–58. Academic Press, New York, 1975.
13. B. Hansson. An analysis of some deontic logics. *Noûs*, 3:373–398, 1969. Reprinted in [**?**, pp 121-147].
14. D. Hume. *A Treatise of Human Nature*. Clarendon Press, Oxford, 2000. David Fate Norton and Mary J. Norton (eds).
15. S.C Levinson. *Pragmatics*. Cambridge University Press, Cambridge, 1983.
16. D.K. Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.
17. D. Makinson. Five faces of minimality. *Studia Logica*, 52:339–379, 1993.
18. D. Makinson. On a fundamental problem in deontic logic. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems*, pages 29–54. IOS Press, Amsterdam, 1999.
19. C. A. Meredith and A. N. Prior. Modal logic with functorial variables and a contingent constant. *Notre Dame J. Formal Logic*, 6(2):99–109, 1965.
20. X. Parent. Cumulativity, identity and time in deontic logic. *Fundam. Inf.*, 48:237–252, 2001.
21. H. Prakken and M. Sergot. Contrary-to-duty obligations. *Studia Logica*, 57:91–115, 1996.
22. H. Prakken and M. Sergot. Dyadic deontic logic and contrary-to-duty obligation. In D. Nute, editor, *Defeasible Deontic Logic*, pages 223–262. Kluwer Academic Publishers, Dordrecht, 1997.
23. M. Sergot. Action and agency in norm-governed multi-agent systems. *Engineering Societies in the Agents World VIII*, pages 1–54, 2008.
24. Y. Shoham. Nonmonotonic reasoning and causation. *Cognitive Science*, 14:213–252, 1990.
25. W. Sinnott-Armstrong. ‘ought’ conversationally implies ‘can’. *Philosophical Review*, 98:249–61, 1984.
26. W. Spohn. An analysis of Hansson’s dyadic deontic logic. *Journal of Philosophical Logic*, 4(2):237–252, 1975.
27. G. H. von Wright. *An Essay in Deontic Logic and the General Theory of Action*. North Holland, Amsterdam, 1968.