CrossMark

Contrary-to-Duty Paradoxes and Counterfactual Deontic Logic

Daniel Rönnedal[1]

Abstract

In this paper, I will discuss some examples of the so-called contrary-to-duty (obligation) paradox, a well-known puzzle in deontic logic. A contrary-to-duty obligation is an obligation telling us what ought to be the case if something forbidden is true, for example: 'If she is guilty, she should confess'. Contrary-to-duty obligations are important in our moral and legal thinking. Therefore, we want to be able to find an adequate symbolisation of such obligations in some logical system, a task that has turned out to be difficult. This is shown by the so-called contrary-to-duty (obligation) paradox. I will investigate and evaluate one kind of solution to this problem that has been suggested in the literature, which I will call the 'counterfactual solution'. I will use some recent systems that combine not only counterfactual logic and deontic logic, but also temporal logic, in my analysis of the paradox. I will argue that the counterfactual solution has many attractive features and that it can give a fairly satisfactory answer to some examples of the contrary-to-duty paradox, but that it nevertheless has some serious problems. The conclusion is that, notwithstanding the many attractive features of the solution, there seem to be other approaches to the paradox that are more promising.

Keywords Contrary-to-duty paradox · Chisholm's paradox · Deontic logic · Contrary-to-duty obligation · Counterfactual logic

1 Introduction

A contrary-to-duty obligation is an obligation telling us what ought to be the case if something forbidden is true. Here are some examples of sentences that often express such obligations: 'If she is guilty, she should confess', 'If he has hurt his friend, he should apologise to her', 'If you are not going to keep your promise to him, you ought to call him', 'If the books are not returned by the due date, you must pay a fine'. We might

✉ Daniel Rönnedal
   daniel.ronnedal@philosophy.su.se

[1]   Department of Philosophy, Stockholm University, Stockholm, Sweden

also say that a contrary-to-duty obligation is a conditional obligation where the condition (in the obligation) is forbidden, or where the condition is fulfilled only if a primary, unconditional obligation is violated. She should not be guilty; but if she is, she should confess. He should not have hurt his friend; but if he did, he should apologise. You should keep your promise to him; but if you are not going to do so, you ought to call him. The books ought to be returned by the due date; but if they are not, you must pay a fine.

Contrary-to-duty obligations turn up in discussions concerning guilt, blame, restoration, reparation, confession, punishment, repentance, retributive justice, compensation, apologies, damage control, etc. Therefore, they are important in our moral and legal thinking. The rationale of a contrary-to-duty obligation is the fact that most of us do neglect our primary duties from time to time and yet it is reasonable to believe that we should make the best of a bad situation, or at least that it matters what we do when this is the case. We want to be able to find an adequate formal representation of such obligations; it has, however, turned out to be difficult to do that. The so-called contrary-to-duty (obligation) paradox, sometimes called the contrary-to-duty imperative paradox, which has been discussed by deontic logicians for more than fifty years now, has clearly shown this. Roderick Chisholm was one of the first philosophers to address this puzzle (Chisholm 1963). Since then many different versions of this problem have been mentioned in the literature (see, e.g., Powers 1967; Åqvist 1967, 2002; Forrester 1984; Prakken and Sergot 1996; and Carmo and Jones 2002, for some examples).

The contrary-to-duty paradox arises when we try to symbolise certain intuitively consistent sets of ordinary language sentences, sets that include at least one contrary-to-duty obligation sentence, by means of ordinary counterparts available in various monadic deontic logics, such as so-called Standard Deontic Logic and similar systems. The formal representations often turn out to be inconsistent, in the sense that it is possible to deduce contradictions from them, or else they might violate some other intuitively plausible condition, for example that the sentences in a representation should be independent of each other.

In this paper, I will discuss some examples of the contrary-to-duty paradox, and I will investigate and evaluate one kind of solution to this puzzle that has been suggested in the literature. I will call this proposal 'the counterfactual solution'. The idea behind this kind of solution is that we can solve the paradox if we combine deontic logic with counterfactual logic. A contrary-to-duty obligation is interpreted as a kind of conditional obligation that involves a counterfactual conditional. Mott (1973) and Niles (1997), for example, are sympathetic to this kind of approach, while Tomberlin (1981) and DeCew (1981) criticise it.

I will not discuss which, if any, contrary-to-duty obligations are true. In this paper, I am primarily interested in the logical form of such obligations and the paradoxes they seem to generate. Before we can answer the question whether a particular contrary-to-duty obligation is true or not, it seems that we must know what we mean by a 'contrary-to-duty obligation'.

In my analysis of the paradox, I will use some recent systems that combine not only counterfactual logic and deontic logic, but also temporal logic (Rönnedal 2016; see the appendix). In those systems, we can avoid some problems with earlier similar counterfactual solutions, and we can make an important distinction between 'afterward' ('backward'), 'beforehand' ('forward') and 'timeless' or 'parallel' contrary-to-duty obligations. In an 'afterward' contrary-to-duty obligation, the content of the obligation

comes after the time of the condition. 'If you do not keep your promise, you ought to apologise (afterwards)' can be used to express such an obligation. The content of the obligation, that you apologise, is supposed to take place after you violate your primary, unconditional obligation to keep your promise, if, in fact, you do not keep it. In a 'beforehand' contrary-to-duty obligation, the content of the obligation comes before the time of the condition. 'If you are not going to keep your promise, you ought to call your friend (beforehand)' can be used to express such an obligation. The content of the obligation, that you call your friend, is supposed to take place before you violate your primary obligation to keep your promise. If, in fact, you will not keep your promise, you ought to call your friend beforehand. In a 'parallel' contrary-to-duty obligation, the content of the obligation is supposed to take place at the same time as the condition. 'If there is a dog, there ought to be a warning sign' can be used to express such an obligation. The content of the obligation, that there is a warning sign, is supposed to be fulfilled at the same time as the primary obligation, that there is no dog, is violated.

I will argue that the counterfactual solution presented in this paper has many attractive features and that it can give a fairly satisfactory answer to some examples of the contrary-to-duty paradox, primarily examples that involve afterward contrary-to-duty obligations, but that it nevertheless has some serious problems, especially with parallel and beforehand contrary-to-duty paradoxes. The conclusion is that, notwithstanding the many attractive features of the solution, there seem to be other approaches to the paradox that are more promising.

The plan of the paper is as follows. In Section 2, I will describe one example of a contrary-to-duty paradox, and in Section 3, I will show how this version can be solved in counterfactual temporal alethic-deontic logic. Section 4 includes several arguments for this solution, and in Section 5, I will consider some problems with the approach. Section 6 contains a summary of the paper and a conclusion.

2 The Contrary-to-Duty (Obligation) Paradox

In this section, I will begin by discussing a particular version of a contrary-to-duty paradox that involves an afterward contrary-to-duty obligation. I will call this version 'the (afterward) promise (contrary-to-duty) paradox'. Then, in later sections, I will mention some other versions.

Scenario I: The (Afterward) Promise (Contrary-to-Duty) Paradox (after Prakken and Sergot 1996) Consider the following scenario. It is Monday and you promise a friend to meet her on Friday to help her with some task. Suppose, further, that you always meet your friend on Saturdays. In this example, the following sentences all seem to be true:

N-CTD

N1   (On Monday it is true that) You ought to keep your promise (and see your friend on Friday).
N2   (On Monday it is true that) It ought to be that if you keep your promise, you do not apologise (when you meet your friend on Saturday).

N3   (On Monday it is true that) If you do not keep your promise (i.e. if you do not see your friend on Friday and help her out), you ought to apologise (when you meet her on Saturday).

N4   (On Monday it is true that) You do not keep your promise (on Friday).

Let N-CTD = {N1, N2, N3, N4}. N3 is a contrary-to-duty obligation (or expresses a contrary-to-duty obligation). If the condition is true, the primary obligation that you should keep your promise (expressed by N1) is violated. Furthermore, it is an afterward contrary-to-duty obligation. If you do not keep your promise on Friday, you ought to apologise afterwards, when you meet your friend on Saturday. N-CTD appears to be consistent, it does not seem possible to deduce any contradiction from this set. Nonetheless, if we try to formalise N-CTD in so-called Standard Deontic Logic, for instance, we immediately encounter some problems. Standard Deontic Logic is a well-known logical system described in most introductions to deontic logic (see e.g. Gabbay, Horty, Parent, van der Meyden and van der Torre (eds.), Gabbay et al. 2013, pp. 36–39). It is basically a normal modal system of the kind KD (Chellas 1980). In Åqvist (2002) this system is called **OK+**. For introductions to deontic logic, see e.g. Hilpinen (1971), (Hilpinen 1981) and Gabbay et al. (2013). Consider the following symbolisation:

SDL-CTD

SDL1 Ok
SDL2 $O(k \rightarrow \neg a)$
SDL3 $\neg k \rightarrow Oa$
SDL4 $\neg k$

O is a sentential operator that takes a sentence as argument and gives a sentence as value. Informally, 'Op' says 'It ought to be (or it should be) the case that (or it is obligatory that) p'. '$\neg$' is the standard negation sign, and '$\rightarrow$' the standard material implication sign, well-known from ordinary propositional logic. In SDL-CTD, 'k' stands for 'You keep your promise (meet your friend on Friday and help her with her task)' and 'a' abbreviates 'You apologise (to your friend for not keeping your promise)'. In this symbolisation, SDL1 is supposed to express a primary obligation and SDL3 a contrary-to-duty obligation telling us what ought to be the case if the primary obligation is violated. However, the set SDL-CTD = {SDL1, SDL2, SDL3, SDL4} is not consistent in Standard Deontic Logic. $O\neg a$ is entailed by SDL1 and SDL2, and from SDL3 and SDL4 we can derive Oa. Hence, we can deduce the following formula from SDL-CTD $Oa \wedge O\neg a$ ('It is obligatory that you apologise and it is obligatory that you do not apologise'), which directly contradicts the so-called axiom D, that is, the schema $\neg(OA \wedge O\neg A)$. ('$\wedge$' is the ordinary symbol for conjunction.) $\neg(OA \wedge O\neg A)$ is included in Standard Deontic Logic (usually as an axiom). Clearly this sentence rules out explicit moral dilemmas. Since N-CTD seems to be consistent, while SDL-CTD is inconsistent, something must be wrong with our formalisation, with Standard Deontic Logic or with our intuitions. In a nutshell, this puzzle is the contrary-to-duty (obligation) paradox.

3 Solutions to the Contrary-to-Duty (Obligation) Paradox: The Counterfactual Solution

There are several possible responses to the contrary-to-duty paradox. The various solutions that have been suggested in the literature can be divided into five categories: quick solutions, e.g. reject some axioms or rules of inference that are necessary to derive the contradiction, try to find some alternative formalisation of N-CTD or reject some of our intuitions about this set; operator solutions that introduce several deontic operators that can be used to symbolise different kinds of unconditional obligations (e.g. Åqvist 1967; Jones and Pörn 1985; Carmo and Jones 2002); connective solutions that interpret the expression 'if, then' in a non-standard way (e.g. Mott 1973; Niles 1997; Bonevac 1998); action or agent solutions that argue that deontic logic should be combined with some kind of action logic (e.g. Castañeda 1977, 1981, 1989; Meyer 1988; Bartha 1993); and temporal solutions, which use some combination of deontic logic and temporal logic to solve the puzzle (e.g. van Eck 1982; Loewer and Belzer 1983; Feldman 1986, 1990; Bartha 1993; Åqvist 2003).

In this section, I will consider the counterfactual solution to the contrary-to-duty paradox and in sections 4 and 5, I discuss some arguments for and against this approach. The idea behind this kind of solution is that we can solve the paradox if we combine deontic logic with counterfactual logic. A contrary-to-duty obligation is interpreted as a kind of conditional obligation that involves a counterfactual conditional. Mott (1973) and Niles (1997), for example, are sympathetic to a view of this kind, while Tomberlin (1981) and DeCew (1981), for instance, criticise it. I will, however, use some recent systems that combine not only counterfactual logic and deontic logic, but also temporal logic, in my analysis of the paradox (Rönnedal 2016; see the appendix). In those systems, we can avoid some problems with earlier similar counterfactual solutions and we can make an important distinction between 'afterward', 'beforehand' and 'parallel' contrary-to-duty obligations. (For some early attempts to combine deontic logic and temporal logic, see e.g. Montague 1968, and Chellas 1969.)

In a counterfactual temporal alethic-deontic logic of the kind introduced by Rönnedal (2016), it is possible to symbolise the concept of a conditional obligation in at least four interesting ways: $(A \,\square\!\rightarrow OB)$, $O(A \,\square\!\rightarrow B)$, $(A \,\square\!\Rightarrow OB)$ and $O(A \,\square\!\Rightarrow B)$. $\square\!\rightarrow$ (and $\square\!\Rightarrow$) is a two-place, sentential operator that takes two sentences as arguments and gives one sentence as value. '$A \,\square\!\rightarrow B$' (and '$A \,\square\!\Rightarrow B$') is often read 'If A were the case, then B would be the case'. (The differences between $\square\!\rightarrow$ and $\square\!\Rightarrow$ are unimportant in this context and I will focus on $\square\!\rightarrow$.) According to the counterfactual solution, contrary-to-duty obligations are conditional obligations of this kind where the condition is forbidden or violates some primary obligation. I will show how counterfactual temporal alethic-deontic logic can be used to avoid the contrary-to-duty paradox.

In this section, I will consider one possible formalisation of N-CTD that seems to be among the most plausible in counterfactual temporal alethic-deontic logic. In sections 4 and 5, I will discuss several other attempts to symbolise some other examples of contrary-to-duty paradoxes. First, however, I will explain some symbols: t_1 (in the formalisation below) refers to the moment on Monday when you make your promise, t_2 refers to the moment on Friday when you should keep your promise and t_3 refers to the moment on Saturday when you should apologise if you do not keep your promise on Friday. R is a temporal operator; 'Rt_1A' says that it is realised at time t_1 (it is true at t_1) that A, etc. N-CTD can now be symbolised in the following way:

CF-CTD

CF1 Rt_1ORt_2k
CF2 $Rt_1O(Rt_2k \,\square\rightarrow Rt_3\neg a)$
CF3 $Rt_1(Rt_2\neg k \,\square\rightarrow Rt_2ORt_3a)$
CF4 $Rt_1Rt_2\neg k$

Note that $Rt_1Rt_2\neg k$ is equivalent to $Rt_2\neg k$. This means that it is true on Monday that it is true on Friday that you do not keep your promise if and only if it is true on Friday that you do not keep your promise. Accordingly, we can use $Rt_2\neg k$ as an alternative symbolisation of N4. Furthermore, note that it might be true on Monday that you will not keep your promise on Friday (in some possible world) even though this is not a settled fact, that is, even though it is not historically necessary. In some possible worlds you will keep your promise on Friday and in some possible worlds you will not. F4 is true at t_1 (i.e. on Monday) in the possible worlds where you do not keep your promise at t_2 (i.e. on Friday).

In counterfactual temporal alethic-deontic logic truth is relativised to world-moment pairs. In other words, a sentence can be true in one possible world ω at a particular time τ even though it is false in some other possible world, say ω', at this time, that is, at τ, or false in this world, that is, in ω, at another time, say τ'. A temporally settled sentence satisfies the following condition: if it is true (in a possible world), it is true at every moment of time (in this possible world); and if it is false (in a possible world), it is false at every moment of time (in this possible world). Some (but not all) sentences are temporally settled. All the sentences in CF-CTD are temporally settled; ORt_2k, $O(Rt_2k \,\square\rightarrow Rt_3\neg a)$ and $O(Rt_2\neg k \,\square\rightarrow Rt_3a)$ are examples of sentences that are not, as their truth values may vary from one moment of time to another (in one and the same possible world).

Let us call the counterfactual temporal alethic-deontic system that includes all basic alethic, deontic and temporal rules, all rules in Tables 3, 4, 5 and 6 and all rules in Table 7 except $T - c7$ in Rönnedal (2016), all alethic rules that correspond to the fact that the alethic accessibility relation is 'reflexive', 'symmetric' and 'transitive', all deontic rules that correspond to the fact that the deontic accessibility relation is 'serial', 'transitive' and 'Euclidean', and all temporal rules that corresponds to the fact that the temporal accessibility relation is 'transitive', 'comparable' and does not branch towards the future or the past, Strong Counterfactual Temporal Alethic-Deontic Logic (**SCTADL**) (see the appendix). **SCTADL** is a very strong counterfactual temporal alethic-deontic system. Basically, all philosophically interesting systems seem to be a part of this logic. Moreover, let CF-CTD = {CF1, CF2, CF3, CF4}. I will show that CF-CTD is consistent in all systems weaker than or deductively equivalent to the system **SCTADL**. To establish this, it is enough to show that CF-CTD is consistent in **SCTADL**. Then it follows that the result holds in all weaker systems too. To show that CF-CTD is consistent in **SCTADL**, it is enough to come up with a model of the right kind where all elements in CF-CTD are true in some possible world at some time. This establishes that the set is satisfiable, and hence, by the soundness results in Rönnedal (2016), that the set is consistent in **SCTADL** and in every weaker system.

Model I Consider the following supplemented counterfactual temporal alethic-deontic model. $W = \{\omega_1, \omega_2, \omega_3, \omega_4\}$; W is a set of possible worlds, ω_1, etc. are possible worlds in W. The expression '$\omega_2 \geq_{\omega_1} \omega_3$' says that ω_2 is at least as similar to ω_1 as is ω_3. We use

'$\omega_2 >_{\omega_1} \omega_3$' to say that ω_2 is more similar to ω_1 than ω_3 is. $\omega_2 >_{\omega_1} \omega_3$ if and only if $\omega_2 \geq_{\omega_1} \omega_3$ and not $\omega_3 \geq_{\omega_1} \omega_2$. We have the following similarity relations between the possible worlds in W: $\omega_1 >_{\omega_1} \omega_2$, $\omega_2 >_{\omega_1} \omega_4$, $\omega_3 >_{\omega_1} \omega_4$, $\omega_2 >_{\omega_2} \omega_1$, $\omega_1 >_{\omega_2} \omega_3$, $\omega_3 >_{\omega_2} \omega_4$, $\omega_3 >_{\omega_3} \omega_4$, $\omega_4 >_{\omega_3} \omega_2$, $\omega_2 >_{\omega_3} \omega_1$, $\omega_4 >_{\omega_4} \omega_3$, $\omega_3 >_{\omega_4} \omega_1$, $\omega_1 >_{\omega_4} \omega_2$. The model satisfies all conditions in Table 2 and all conditions except $C - c7$ in Table 1 in Rönnedal (2016) (see the appendix). The alethic accessibility relation is an 'equivalence' relation (i.e. it is 'reflexive', 'symmetric', and 'transitive' (at every moment in time), etc.). The deontic accessibility relation is 'serial', 'transitive' and 'Euclidean' (and even 'functional') (at every moment in time). $T = \{\tau_1, \tau_2, \tau_3\}$, where T is a set of moments of time, and τ_1, etc., are particular moments of time in T. $\tau_1 < \tau_2 < \tau_3$. '$\tau_1 < \tau_2$' says that τ_1 is before τ_2. The temporal relation is 'transitive', 'comparable', and it does not branch towards the future or the past. At τ_1, ω_1 is deontically accessible from every possible world (and no other world is deontically accessible from any world). ω_1 is deontically accessible from ω_1 and ω_2, and ω_3 is deontically accessible from ω_3 and ω_4 at τ_2. No other possible world is deontically accessible from any world at τ_2. At τ_3, every possible world is deontically accessible from itself. No possible world is deontically accessible from any other possible world at τ_3. At τ_1, all possible worlds are alethically accessible from all possible worlds. At τ_2, ω_2 is alethically accessible from ω_1 and ω_1 is alethically accessible from ω_2. At τ_2, ω_3 can see ω_4 alethically and ω_4 can see ω_3 alethically. At all times every possible world is alethically accessible to itself. $v(t_1) = \tau_1$, $v(t_2) = \tau_2$, and $v(t_3) = \tau_3$. k is true in ω_1 and ω_2 at τ_2 and k is false in ω_3 and ω_4 at τ_2. a is false in ω_1 and ω_4 at τ_3 and a is true in ω_2 and ω_3 at τ_3. For our purposes, we do not need any further information about this model. Model I is a model of the right kind.

In ω_3 at τ_1, all sentences in CF-CTD are true. Hence, by the soundness results in Rönnedal (2016), the set is consistent in **SCTADL** and in every weaker system. Therefore, it appears to be the case that we can solve the contrary-to-duty paradox in counterfactual temporal alethic-deontic logic, at least the promise paradox. CF-CTD is not inconsistent in any philosophically interesting counterfactual temporal alethic-deontic system.

4 Arguments for the Counterfactual Solution

In this section, I will discuss some reasons why the counterfactual solution is attractive. Together these reasons provide a strong, although not conclusive, case for this approach.[1]

---

[1] Carmo and Jones (2002) suggest that a plausible solution to the contrary-to-duty paradox should meet certain conditions: (i) consistency; (ii) logical independence; (iii) applicability to (at least apparently) timeless and actionless contrary-to-duty examples; (iv) analogous logical structures for conditional sentences; (v) capacity to derive actual obligations; (vi) capacity to derive ideal obligations; (vii) capacity to represent the fact that a violation of an obligation has occurred; (viii) capacity to avoid the pragmatic oddity; (ix) the assignment of logical form should be independent. In this section, we will see that the counterfactual solution meets all of these conditions except (iv) (and perhaps the first part of (iii); see Section 5.2), and that it has some other attractive features. The solution in Carmo and Jones (2002) meets all requirements they mention. But this solution has other problems that I cannot discuss in this paper. It is unlikely that there is a solution that meets all the conditions philosophers have thought a solution to the paradox should meet. We probably have to give up some of our pre-philosophical intuitions. This is why the paradox is so profound.

Reason (1): CF-CTD is consistent We have shown above that CF-CTD is consistent (in **SCTADL** and in every weaker system). It does not seem to be possible to deduce any contradiction from N-CTD. Hence, we want our symbolisation of this set to be consistent. In this respect, CF-CTD is an intuitively plausible formalisation of N-CTD.

Reason (2): CF-CTD is non-redundant A set of sentences is non-redundant if and only if no sentence in this set follows from the rest. N-CTD appears to be non-redundant. Consequently, we want our symbolisation of N-CTD to be non-redundant. In certain monomodal deontic systems, for instance Standard Deontic Logic, we can try to solve the contrary-to-duty paradox by finding some other formalisation than SDL-CTD of the sentences in N-CTD. Instead of SDL2 we can use $k \rightarrow O\neg a$ and instead of SDL3 we can use $O(\neg k \rightarrow a)$. If we employ these formulas, we obtain three consistent alternative symbolisations of N-CTD. Anyhow, $O(\neg k \rightarrow a)$ follows from Ok in every so-called normal deontic system, including Standard Deontic Logic, and $k \rightarrow O\neg a$ follows from $\neg k$ by propositional logic. So, these alternatives are not non-redundant. Intuitively, N3 does not seem to be deducible from N1, and N2 does not seem to follow from N4. In counterfactual temporal alethic-deontic logic, we can avoid this problem. All sentences in CF-CTD are independent of each other in **SCTADL** and in every weaker system. To establish this claim, it is enough to show that the elements of CF-CTD are independent of each other in **SCTADL**. Then it follows that the result also holds for all weaker systems. To prove this, I will establish the following four propositions.

Proposition 1 CF1 is not deducible from {CF2, CF3, CF4} in **SCTADL**. To prove this proposition, it is sufficient to come up with a model M of the right kind, a world ω in M and a time τ in M such that all members of {CF2, CF3, CF4} are true in ω at τ and CF1 false in this world at this time. Then the proposition follows by the soundness results in Rönnedal (2016). Consider the following model.

Model II Model II is exactly like Model I except that the deontic accessibility relation is modified in the following way. At τ_1, ω_4 is deontically accessible from every possible world. No other possible world is deontically accessible from any possible world at τ_1. All members of {CF2, CF3, CF4} are true and CF1 false in world ω_4 at time τ_1. Consequently, CF1 is not deducible from {CF2, CF3, CF4} in **SCTADL** (or any weaker system).

Proposition 2 CF2 is not deducible from {CF1, CF3, CF4} in **SCTADL**. The following model proves this proposition.

Model III Model III is exactly like Model I except that a is true in ω_1 at τ_3. All members of {CF1, CF3, CF4} are true and CF2 false in world ω_3 at time τ_1. Therefore, CF2 is not deducible from {CF1, CF3, CF4} in **SCTADL** (or any weaker system).

Proposition 3 CF3 is not derivable from {CF1, CF2, CF4} in **SCTADL**. To prove this claim we use the following model.

Model IV Model IV is exactly like Model I except that a is false in ω_3 at time τ_3. All members of {CF1, CF2, CF4} are true and CF3 false in world ω_1 at time τ_1. Hence, CF3 is not deducible from {CF1, CF2, CF4} in **SCTADL** (or any weaker system).

Proposition 4 CF4 is not derivable from {CF1, CF2, CF3} in **SCTADL**. Consider Model I: in world ω_2 at time τ_1 all members of {CF1, CF2, CF3} are true and CF4 false. Hence, CF4 is not derivable from {CF1, CF2, CF3} in **SCTADL** (or any weaker system), by the soundness of **SCTADL**.[2]

Reason (3): CF-CTD is dilemma free Standard Deontic Logic minus the axiom D, $\neg(OA \land O\neg A)$, is the minimal so-called normal deontic logic K. SDL-CTD does not entail any contradiction in K. Hence, it is possible to avoid the contrary-to-duty paradox in ordinary monomodal deontic logic if we give up this axiom. There might be independent reasons for rejecting D. If there are, or could be, genuine moral dilemmas, D cannot in general be true. But even if there are or could be genuine moral dilemmas (which is debatable), rejecting D does not seem to be a good solution to the contrary-to-duty paradox for several reasons. Firstly, even if we reject axiom D, it is counterintuitive to claim that a dilemma follows from N-CTD. We can still deduce the sentence $Oa \land O\neg a$ from SDL-CTD in every normal deontic system, which says that it is obligatory that you apologise and it is obligatory that you do not apologise. And this proposition does not seem to follow from N-CTD. Secondly, if there are any moral dilemmas of this kind, we can derive the claim that everything is both obligatory and forbidden in every normal deontic system, which is absurd (see Reason 4 below). Thirdly, such a solution might still have problems with the so-called pragmatic oddity (see Reason 5 below).

The solution in counterfactual temporal alethic-deontic logic is dilemma free. The sentence $Rt_1ORt_3\neg a$ is derivable from CF1 and CF2 (see Reason 10 below) and from CF3 and CF4 we can deduce the formula Rt_2ORt_3a (see Reason 11 below) (in some systems). Thus, it is possible to derive the following sentence from CF-CTD: $Rt_1ORt_3\neg a \land Rt_2ORt_3a$ (in certain systems). '$Rt_1ORt_3\neg a$' says 'On Monday [when you have not yet broken your promise] it ought to be the case that you do not apologise on Saturday', and 'Rt_2ORt_3a' says 'On Friday [when you have broken your promise] it ought to be the case that you apologise on Saturday'. Nonetheless, ORt_3a and $ORt_3\neg a$ are not true at the same time. Neither $Rt_1ORt_3\neg a \land Rt_1ORt_3a$ nor $Rt_2ORt_3\neg a \land Rt_2ORt_3a$ is deducible from CF-CTD in any interesting counterfactual temporal alethic-deontic system. So, this is not a moral dilemma. Since N-CTD seems to be dilemma free, we want our formalisation of N-CTD to be dilemma free too; and CF-CTD is, as we have seen, dilemma free. This is one good reason to be attracted to the counterfactual solution.

Reason (4): It is not possible to derive the proposition that everything is both obligatory and forbidden from CF-CTD If there is at least one moral dilemma, we can derive the conclusion that everything is both obligatory and forbidden in every normal deontic logic (even without the axiom D). Since FA ('It is forbidden that A') is equivalent to $O\neg A$, this follows from the fact that $Oa \land O\neg a$ entails Or for any r in every normal deontic logic. This is clearly absurd. N-CTD does not appear to entail that everything is both obligatory and forbidden. Accordingly, we do not want our symbolisation to entail this. In counterfactual temporal alethic-deontic logic we can avoid this problem. Consider Model I above. Every sentence in CF-CTD is true in ω_3 at

---

[2] The idea that a solution to the contrary-to-duty paradoxes should be consistent and non-redundant is old, see for example Åqvist (1967).

time τ_1 in this model. For all that, $Rt_1ORt_3\neg a$ is true while $Rt_1O\neg Rt_3\neg a$ and $Rt_1FRt_3\neg a$ are false in this world at this time.

Reason (5): CF-CTD does not entail the so-called pragmatic oddity It does not seem to follow from N-CTD that you should keep your promise *and* apologise (for not keeping your promise). Accordingly, we do not want our symbolisation of N-CTD to entail this counterintuitive conclusion or anything similar to it. Despite that, $O(k \wedge a)$ is derivable from SDL-CTD in every so-called normal deontic logic (with or without the axiom D), and '$O(k \wedge a)$' says that it is obligatory that you keep your promise *and* apologise (for not keeping your promise). Even if this claim is not inconsistent, it is certainly very odd. This is the so-called pragmatic oddity, which is a problem for many possible solutions to the contrary-to-duty paradox (Prakken and Sergot 1996). In counterfactual temporal alethic-deontic logic, we can avoid this puzzle. Neither $O(Rt_2k \wedge Rt_3a)$, $Rt_1O(Rt_2k \wedge Rt_3a)$ nor $Rt_2O(Rt_2k \wedge Rt_3a)$ is deducible from CF-CTD in any interesting counterfactual temporal alethic-deontic system. Consider Model I above. In ω_3 at τ_1 all sentences in CF-CTD are true, but neither $O(Rt_2k \wedge Rt_3a)$, $Rt_1O(Rt_2k \wedge Rt_3a)$ nor $Rt_2O(Rt_2k \wedge Rt_3a)$ holds in this world at this time. This proves our claim.

Reason (6): The counterfactual solution is applicable to (at least apparently) actionless and agentless contrary-to-duty examples If we combine deontic logic with some kind of action logic, for example some kind of *Stit* ('Seeing to it') logic (Bartha 1993), or dynamic logic (Meyer 1988), it might be possible to solve some contrary-to-duty paradoxes. However, there also seem to be examples of contrary-to-duty paradoxes that involve action- and agentless contrary-to-duty obligations. It appears to be difficult to solve these paradoxes in such systems. (For more on *Stit* logic, see Horty 2001, and Belnap et al. 2001).

Scenario II: Contrary-to-Duty Paradoxes Involving (apparently) action- and/or agentless contrary-to-duty obligations Consider the following scenario. At t_1 (some arbitrary point in time when there is no fire in the house), it ought to be the case that there is no fire in the house at t_2 (some time later than t_1). (At t_1) It ought to be that if there is no fire in the house (at t_2), then it is not the case that the sprinkler goes off (at t_3, a point in time as soon as possible after t_2). (At t_1) If there is a fire in the house (at t_2), then (at t_2) the sprinkler ought to go off (at t_3). Moreover, there is a fire in the house (at t_2). In this example, all of the following claims seem to be true:

N-ACTD

AN1   (At t_1) There ought to be no fire in the house (at t_2).

AN2   (At t_1) It ought to be that if there is no fire in the house (at t_2), then it is not the case that the sprinkler goes off (at t_3).

AN3   (At t_1) If there is a fire in the house (at t_2) then (at t_2) the sprinkler ought to go off (at t_3).

AN4   (At t_1 it is the case that at t_2) There is a fire in the house.

AN1 tells us what ought to be the case unconditionally, and AN3 expresses a contrary-to-duty obligation. The condition in AN3 is fulfilled only if the primary

obligation expressed by AN1 is violated. But AN3 does not seem to involve any action or particular agent in any ordinary sense. The scenario is an example of an action- and agentless contrary-to-duty paradox. AN3 says something about what ought to be the case if the world is not as it ought to be according to AN1. It is difficult to find a plausible symbolisation of N-ACTD and similar paradoxes in dynamic deontic logic or in *Stit* logic.

But for all that, in counterfactual temporal alethic-deontic logic, we have no trouble symbolising (apparently) action- and agentless contrary-to-duty obligations. The logical form of the sentences in N-ACTD exactly parallels the logical form of the sentences in N-CTD. Hence, we can solve contrary-to-duty paradoxes of this kind in exactly the same way as we solved our original paradox.[3]

Reason (7): We do not have to postulate any primitive conditional obligations In counterfactual temporal alethic-deontic logic, a conditional obligation can be expressed by a combination of a counterfactual conditional and an ordinary (unconditional) obligation. We do not have to introduce any new primitive dyadic deontic operators. According to various dyadic and temporal dyadic deontic solutions, we need some new primitive dyadic deontic operator (e.g. van Eck 1982; Loewer and Belzer 1983; Feldman 1986, 1990; Åqvist 2003). (For more on dyadic deontic logic, see e.g. Rescher 1958; von Wright 1964; Danielsson 1968; Hansson 1969; van Fraassen 1972, 1973; Lewis 1974; von Kutschera 1974; Spohn 1975; Cox 1978; Åqvist 1984, 1987, 2002; Åqvist and Hoepelman 1981, Goble 2003, 2004; Parent 2008, 2010.) Some philosophers have been critical of this approach (e.g. Bonevac 1998). They think a proper theory of conditional obligations, including contrary-to-duty obligations, will be the product of two separate components: a theory of the conditional and a theory of obligation. The counterfactual solution satisfies this 'requirement'. (For some other criticisms, see e.g. Tomberlin 1989, 1989b.)

Reason (8): We can express the idea that an obligation has been violated It has been suggested that the contrary-to-duty paradox can be solved by using some kind of nonmonotonic logic (see e.g. Bonevac 1998). However, it is not obvious that we can explain the difference between violation and defeat in such a logic. If you will not see your friend and help her, the primary, unconditional obligation to keep your promise will be violated. It is not the case that this obligation is defeated, overridden or cancelled. The conditional norms in N-CTD do not cancel each other out. Nor is it the case that one of the norms defeat or override the other. Smith (1994) and Prakken and Sergot (1996) stress the difference between violation and defeat.

In counterfactual temporal alethic-deontic logic, it is easy to express the idea that an obligation will be violated, or has been violated. At t_1, CF4 $[Rt_1Rt_2\neg k]$—for example—expresses the idea that the unconditional norm CF1 $[Rt_1ORt_2k]$ will be violated. The fact that a norm will be violated does not entail that this is necessary, settled or determined. At t_2, on the other hand, we can say that it is settled or

---

[3] Some might argue that there are no agentless obligations. In this paper, I do not make a sharp distinction between obligations and 'oughts'. But we could make such a distinction, and then we could speak about agentless 'oughts' instead of agentless obligations. According to some deontic logicians, deontic statements *must* take 'agentives' as their complements (Belnap et al. 2001, p. 13). But this view is not shared by everyone (see e.g. McNamara 2004). To me, the sentences in N-ACTD clearly seem to be meaningful.

historically necessary that you do not keep your promise (in the worlds where you in fact do not keep it). Then, we might want to say that it is no longer obligatory that you keep your promise. In spite of that, the sentence Rt_1ORt_2k is still true at t_2 (in the worlds where you do not keep your promise), that is, it is still true that you should have kept your promise. So, when you do not keep your promise, you violate this 'earlier' duty.

Reason (9): We can symbolise higher order contrary-to-duty obligations There are contrary-to-duty obligations of a 'higher order' or 'degree'. This is illustrated by the following scenario:

Scenario III: The Higher Order Contrary-to-Duty Paradox Scenario III is similar to Scenario I. Although, we now suppose that you do not meet your friend on Saturday and apologise. So, you do not only violate the primary obligation to keep your promise, but also the contrary-to-duty obligation to apologise on Saturday given that you do not keep your promise. In this case, we can assume that there is a 'higher order' or 'degree' contrary-to-contrary-to-duty obligation to apologise (on Sunday when you do in fact meet your friend) given that you neither keep your promise nor apologise on Saturday. In this scenario all of the following sentences seem to be true:

N-HCTD

HN1   (On Monday it is true that) You ought to keep your promise (and see your friend on Friday).
HN2   (On Monday it is true that) It ought to be that if you keep your promise, you do not apologise (on Saturday).
HN3   (On Monday it is true that) It ought to be that if you keep your promise, you do not apologise (on Sunday).
HN4   (On Monday it is true that) If you do not keep your promise (i.e. if you will not see your friend on Friday and help her out), then (on Friday) you ought to apologise (on Saturday).
HN5   (On Monday it is true that) You do not keep your promise (on Friday).
HN6   (On Monday it is true that) If you do not keep your promise (on Friday) and you do not apologise (on Saturday), then (on Saturday) you ought to apologise (on Sunday).
HN7   (On Monday it is true that) You do not apologise (on Saturday).

Let N-HCTD = {HN1,. .., HN7}. HN4 is an ordinary, first-order or first-degree contrary-to-duty obligation that tells us what ought to be the case if the primary obligation expressed by HN1 is violated. HN6 expresses a contrary-to-contrary to duty obligation, a second-order or second-degree contrary-to-duty obligation. The condition in this obligation is satisfied only if the primary obligation expressed by HN1 is violated and the first-order contrary-to-duty obligation to apologise (on Saturday) is violated. A plausible solution to the contrary-to-duty paradox should be able to deal with higher-order contrary-to-duty obligations as well as ordinary first-degree contrary-to-duty obligations. In

our counterfactual temporal alethic-deontic systems, we can symbolise such higher-order contrary-to-duty obligations. N-HCTD can, for example, be formalised in the following way in counterfactual temporal alethic-deontic logic:

F-HCTD

HF1 Rt_1ORt_2k
HF2 $Rt_1O(Rt_2k \;\square\!\rightarrow Rt_3\neg a)$
HF3 $Rt_1O(Rt_2k \;\square\!\rightarrow Rt_4\neg a)$
HF4 $Rt_1(Rt_2\neg k \;\square\!\rightarrow Rt_2ORt_3a)$
HF5 $Rt_1Rt_2\neg k \; [Rt_2\neg k]$
HF6 $Rt_1((Rt_2\neg k \wedge Rt_3\neg a) \;\square\!\rightarrow Rt_3ORt_4a)$
HF7 $Rt_1Rt_3\neg a \; [Rt_3\neg a]$

Let F-HCTD = {HF1,..., HF7}. N-HCTD seems to be consistent, non-redundant, etc. Hence, we want our symbolisation of this set to have these properties too. Not all solutions to the contrary-to-duty paradox can solve such 'higher order' contrary-to-duty paradoxes. Yet, F-HCTD is consistent, non-redundant, etc. This is a good reason to accept the counterfactual solution.

Reason (10): We can derive 'ideal' obligations from CF-CTD N1 and N2 in N-CTD seem to entail that you ought not to apologise. You ought to keep your promise, and it ought to be that if you keep your promise, then you do not apologise (for not keeping your promise). So, ideally you ought not to apologise. We want the symbolisation of N-CTD to reflect this intuition. In many counterfactual temporal alethic-deontic systems, for instance **SCTADL**, $Rt_1ORt_3\neg a$ is deducible from CF1 [Rt_1ORt_2k] and CF2 [$Rt_1O(Rt_2k \;\square\!\rightarrow Rt_3\neg a)$]. The tableau below proves this.

I will use two derived rules in the deduction. These rules are also employed in the derivation in Reason 11 below. According to the first derived rule, $DR1$, we may add $\neg A$, w_it to any open branch in a tree that includes $\neg RtA$, w_it_j. This rule holds in every system. According to the second derived rule, $DR2$, we may add $A \rightarrow B$, w_it_j to any open branch in a tree that contains $A \;\square\!\rightarrow B$, w_it_j. $DR2$ can be deduced in every system that includes the rule $T - c5$ (and CUT). MP is the derived rule 'Modus Ponens'. (For more information on the rules that we use in our deductions, see Rönnedal 2016, and the appendix.)

(1) Rt_1ORt_2k, w_0t_0
(2) $Rt_1O(Rt_2k \;\square\!\rightarrow Rt_3\neg a)$, w_0t_0
(3) $\neg Rt_1ORt_3\neg a$, w_0t_0
(4) $\neg ORt_3\neg a$, w_0t_1 [3, $DR1$]
(5) $P\neg Rt_3\neg a$, w_0t_1 [4, $\neg O$]
(6) $sw_0w_1t_1$ [5, P]
(7) $\neg Rt_3\neg a$, w_1t_1 [5, P]
(8) $\neg\neg a$, w_1t_3 [7, $DR1$]
(9) ORt_2k, w_0t_1 [1, Rt]
(10) $O(Rt_2k \;\square\!\rightarrow Rt_3\neg a)$, w_0t_1 [2, Rt]
(11) Rt_2k, w_1t_1 [6, 9, O]

(12) $Rt_2k \,\square\!\rightarrow Rt_3\neg a,\ w_1t_1$ [6, 10, O]
(13) $Rt_2k \rightarrow Rt_3\neg a,\ w_1t_1$ [12, $DR2$]
(14) $Rt_3\neg a,\ w_1t_1$ [11, 13, MP]
(15) $\neg a,\ w_1t_3$ [14, Rt]
(16) $*$ [8, 15]

Informally, '$Rt_1ORt_3\neg a$' says that it is true at t_1 (i.e. on Monday) that it ought to be the case that you will not apologise on Saturday when you meet your friend. Even so, $Rt_2ORt_3\neg a$ is not derivable from CF1 and CF2 (see Reason 3 above). On Friday, when you have broken your promise, and when it is no longer historically possible for you to keep your promise, then it is not obligatory that you do not apologise on Saturday. On the contrary, on Friday, it is obligatory that you apologise when you meet your friend on Saturday (see Reason 11). Nevertheless, it is reasonable to claim that it is true on Monday that it ought to be the case that you do not apologise on Saturday. For on Monday it is not a settled fact that you will not keep your promise; on Monday, it is still possible for you to keep your promise, which you ought to do. These conclusions correspond well with our intuitions about Scenario I.

Reason (11): We can derive 'actual' obligations from CF-CTD Ideally, you ought to keep your promise. But if you do not keep your promise, you ought to apologise. And in fact, you do not keep your promise. Therefore, you ought to apologise. In other words, N3 and N4 seem to entail that you ought to apologise. We want our symbolisation of N-CTD to reflect this intuition. According to Rt_2ORt_3a, it is true at t_2 (i.e. on Friday) when you have broken your promise to your friend, that it ought to be the case that you apologise to your friend on Saturday when you meet her. We will now show that Rt_2ORt_3a is derivable from $Rt_1(Rt_2\neg k \,\square\!\rightarrow Rt_2ORt_3a)$ and $Rt_1Rt_2\neg k$.

(1) $Rt_1(Rt_2\neg k \,\square\!\rightarrow Rt_2ORt_3a),\ w_0t_0$
(2) $Rt_1Rt_2\neg k,\ w_0t_0$
(3) $\neg Rt_2ORt_3a,\ w_0t_0$
(4) $\neg ORt_3a,\ w_0t_2$ [3, $DR1$]
(5) $P\neg Rt_3a,\ w_0t_2$ [4, $\neg O$]
(6) $sw_0w_1t_2$ [5, P]
(7) $\neg Rt_3a,\ w_1t_2$ [5, P]
(8) $\neg a,\ w_1t_3$ [7, $DR1$]
(9) $Rt_2\neg k \,\square\!\rightarrow Rt_2ORt_3a,\ w_0t_1$ [1, Rt]
(10) $Rt_2\neg k,\ w_0t_1$ [2, Rt]
(11) $Rt_2\neg k \rightarrow Rt_2ORt_3a,\ w_0t_1$ [9, $DR2$]
(12) $Rt_2ORt_3a,\ w_0t_1$ [10, 11, MP]
(13) $ORt_3a,\ w_0t_2$ [12, Rt]
(14) $Rt_3a,\ w_1t_2$ [6, 13, O]
(15) $a,\ w_1t_3$ [14, Rt]
(16) $*$ [8, 15]

Note that Rt_1ORt_3a is not derivable from CF3 and CF4 (see Reason 3). According to Rt_1ORt_3a, it is true at t_1 (i.e. on Monday) that you should apologise to you friend on Saturday when you meet her. But on Monday it is not yet a settled fact that you will not

keep your promise to your friend, on Monday it is still open to you to keep your promise. Accordingly, it is not true on Monday that you should apologise on Saturday. Since it is true on Monday that you ought to keep your promise, and it ought to be that if you keep your promise then you do not apologise, it follows that it is true on Monday that it ought to be the case that you do not apologise on Saturday (see Reason 10). These facts correspond well with our intuitions about Scenario I.

Reason (12): We can avoid the so-called dilemma of commitment and detachment (Factual) Detachment is an inference pattern that allows us to infer or detach an unconditional obligation from a conditional obligation, given that the condition obtains. Thus, if detachment holds for the conditional (contrary-to-duty) obligation that you should apologise if you do not keep your promise (if detachment is possible), then if it is in fact true that you do not keep your promise, we can derive the unconditional obligation that you should apologise.

According to the so-called dilemma of commitment and detachment (van Eck 1982, p. 263; see also Greenspan, 1975): (1) Detachment should be possible, for we cannot take seriously a conditional obligation if we cannot detach an unconditional obligation from it; and (2) Detachment should not be possible, for if detachment is possible, the following kind of situation would be inconsistent—A, it ought to be the case that B given that A; and C, it ought to be the case that not-B given C. But, such a situation is not necessarily inconsistent.[4]

This dilemma appears to be a problem for many different solutions to the contrary-to-duty paradox, for example solutions that use some kind of pure dyadic deontic logic. In pure dyadic deontic logic, we cannot derive the unconditional obligation that it is obligatory that A (OA) from the dyadic obligation that it is obligatory that A given B ($O(A/B)$, $O[B]A$) and B. But if this is true, how can we take such conditional obligations seriously? In counterfactual temporal alethic-deontic logic we can avoid this dilemma. From A $\square \rightarrow OB$ and A we can detach OB (in most interesting systems, for example **SCTADL**). However, the following set of formulas is not necessarily inconsistent: $\{A, O(A \ \square \rightarrow B), C, O(C \ \square \rightarrow \neg B)\}$. We have seen above that Rt_2ORt_3a, but not Rt_1ORt_3a, is detachable from $Rt_1Rt_2\neg k$ and $Rt_1(Rt_2\neg k \ \square \rightarrow Rt_2ORt_3a)$ (in certain systems). These conclusions correspond well with our intuitions.

Reason (13): The structure of the sentences in CF-CTD reflects the structure of the sentences in N-CTD In N2 (It ought to be that if you keep your promise, you do not apologise.), 'ought' has 'wide scope' and in N3 (If you do not keep your promise, you ought to apologise.) 'ought' has narrow scope. This is reflected in the formal sentences CF2 and CF3. In CF2 [$Rt_1O(Rt_2k \ \square \rightarrow Rt_3\neg a)$], '$O$' has wide scope and in CF3 [$Rt_1(Rt_2\neg k \ \square \rightarrow Rt_2ORt_3a)$] '$O$' has narrow scope. Some solutions to the paradox do not respect this difference in structure between N2 and N3.

---

[4] Some philosophers think that detachment should hold, for example Castañeda (1989), while some think it should not hold, at least not without restrictions, for example Greenspan (1975), DeCew (1981) and Feldman (1989).

Reason (14): We do not have to postulate several different unconditional obligation operators Some philosophers have suggested that we should solve the contrary-to-duty paradox by introducing different kinds of obligations, symbolised by distinct obligation operators (e.g. Åqvist 1967; Jones and Pörn 1985; and Carmo and Jones 2002). Assume, for example, that there are two obligation operators O_1 and O_2 that represent ideal and actual obligations, respectively. Then it might be possible to derive the formula $O_1 \neg a \land O_2 a$ from CF-CTD instead of $Oa \land O \neg a$. But $O_1 \neg a \land O_2 a$ is not inconsistent with the axiom D; '$O_1 \neg a \land O_2 a$' says that it is 'ideally-obligatory' that you do not apologise and it is 'actually-obligatory' that you apologise. Still, this solution is problematic, since the derived 'ideal' obligation not to apologise does not seem to be of another kind compared to the derived 'actual' obligation to apologise.

In counterfactual temporal alethic-deontic logic, we do not have to postulate several different kinds of unconditional obligations. The unconditional obligation to not apologise is the same kind of obligation as the derived unconditional obligation to apologise. The only difference is that they hold at different times; the 'ideal' obligation not to apologise holds at τ_1 (on Monday) and the 'actual' obligation to apologise holds at τ_2 (on Friday).

Reason (15): The assignment of logical form to each of the norms in the representation of a contrary-to-duty scenario is independent of the other norms in it According to Carmo and Jones (2002), the assignment of logical form to each of the norms in the representation of a contrary-to-duty scenario should be independent of the other norms in it. The counterfactual solution meets this requirement. Not all solutions do that (see Carmo and Jones 2002).

This ends our discussion of reasons for the counterfactual solution. In light of these arguments, the counterfactual solution seems to be among the most plausible so far suggested in the literature. It appears to be the case that it can solve at least contrary-to-duty paradoxes that involve afterward contrary-to-duty obligations. Despite that, it also has some serious problems. In the next section, I will consider some arguments against this solution. (For more on some problems, see DeCew 1981, and Tomberlin 1981, and for some responses, see Niles 1997.)

5 Problems with the Counterfactual Solution

In this section, I will consider three arguments against the counterfactual solution. According to the first one, the solution introduced in this paper does not add anything new to the discussion. According to the second argument, the counterfactual solution cannot handle timeless (or parallel) contrary-to-duty paradoxes, and according to the third one, it cannot solve beforehand contrary-to-duty paradoxes. It seems to me that the first argument can be answered. The solution in this paper *does* add something new to the discussion, and this is a good reason why it is interesting to investigate it. The second argument is more problematic, even though I do not think it is conclusive (see the reply in Section 5.2.5). The third is probably the most disturbing. Moreover, it suggests that there are other solutions to the paradox that are better than the counterfactual solution, all things considered, notwithstanding the fact that the counterfactual solution is very attractive and that it appears to be the case that it can solve afterward contrary-to-duty paradoxes.

5.1 Argument 1: The Solution in this Paper Does Not Add Anything New to the Discussion

According to the first argument, the counterfactual solution in this paper does not add anything new to the discussion. The argument consists of two propositions. According to the first, the solution does not add anything because we can already solve the contrary-to-duty paradox in counterfactual deontic logic without a temporal part. According to the second, it does not add anything since we can solve the contrary-to-duty puzzle in a temporal deontic logic without any counterfactual part. I will now argue that both of these claims are problematic.

In pure counterfactual deontic logic, we can symbolise N-CTD in the following way: $\{Ok, O(k \ \square \rightarrow \neg a), \neg k \ \square \rightarrow Oa, \neg k\}$. Nevertheless, in most systems, this set is inconsistent, whereas CF-CTD is consistent. This problem can be avoided by using the following formalisation instead: $\{Ok, k \ \square \rightarrow O\neg a, \neg k \ \square \rightarrow Oa, \neg k\}$. But there are other problems with this symbolisation. We can, for example, not derive the 'ideal' obligation that you should not apologise from this set. Furthermore, both of these solutions generate the pragmatic oddity. In other words, both sets entail the following formula, $O(k \wedge a)$, which says that it is obligatory that you keep your promise *and* apologise (for not keeping your promise). This is surely counterintuitive. Moreover, it is difficult to find any other plausible symbolisation of N-CTD in pure counterfactual deontic logic without any temporal part. (See DeCew 1981, for some further critique.) The counterfactual temporal solution does not have any of these problems. So, it is clearly a step forward. It is better than similar solutions in pure counterfactual deontic logic without a temporal part, at least in some important respects. In the light of all the arguments for the solution, it is clearly worth investigating it.

According to the second proposition in this argument, we do not really need the counterfactual in our symbolisation. We can solve afterward contrary-to-duty paradoxes already in temporal deontic logic. We can use the following symbolisation instead: $\{Rt_1ORt_2k, Rt_1O(Rt_2k \rightarrow Rt_3\neg a), Rt_1(Rt_2\neg k \rightarrow Rt_2ORt_3a), Rt_1Rt_2\neg k\}$. It seems that this solution has all the advantages that the counterfactual solution has. The set is consistent, non-redundant, etc. (in most, perhaps all, interesting logics). So, we do not really need to include any counterfactual part in our systems. However, even if it is true that this set is consistent, non-redundant, etc., there are other arguments against this kind of solution. Suppose that instead of $Rt_1Rt_2\neg k$ we have Rt_1Rt_2k. Then it follows that $Rt_1(Rt_2\neg k \rightarrow Rt_2ORt_3\neg a)$ and, in fact, that $Rt_1(Rt_2\neg k \rightarrow Rt_2ORt_3A)$ for any A. $Rt_1(Rt_2\neg k \rightarrow Rt_2ORt_3a)$ is supposed to express a contrary-to-duty obligation. But if this is true, it seems that $Rt_1(Rt_2\neg k \rightarrow Rt_2ORt_3\neg a)$ and $Rt_1(Rt_2\neg k \rightarrow Rt_2ORt_3A)$ also express contrary-to-duty obligations. This is counterintuitive. It is implausible to claim that the proposition that you keep your promise on Friday entails the following contrary-to-duty obligation: if you do not keep your promise (on Friday), then you ought *not* to apologise (on Saturday). In fact, if $Rt_1(Rt_2\neg k \rightarrow Rt_2ORt_3A)$ expresses a contrary-to-duty obligation, we can derive a whole set of problematic conclusions, for example, that if you do not keep your promise (on Friday), then you ought to lie, steal, cheat, rape, etc. (on Saturday). This is

surely counterintuitive. The counterfactual solution does not have this problem. From Rt_1Rt_2k, it does not follow that $Rt_1(Rt_2\neg k \ \Box \rightarrow Rt_2ORt_3A)$. In this respect, the counterfactual solution is better than the 'simple' temporal solution.

In conclusion, the first argument does not appear to succeed. The counterfactual solution in this paper *does* add something new and interesting to the debate. Moreover, it seems to be a fairly plausible solution to afterward contrary-to-duty paradoxes. So, let us turn to the second, more problematic argument.

5.2 Argument 2: The Counterfactual Solution cannot Handle Timeless (or Parallel) Contrary-to-Duty Paradoxes

There seems to be timeless (or parallel) contrary-to-duty paradoxes. In a timeless (or parallel) contrary-to-duty paradox, all obligations appear, in some sense, to be in force simultaneously and both the antecedent and consequent in the contrary-to-duty obligation seem to 'refer' to the same time (if indeed they refer to any time at all). Paradoxes of this kind cannot be solved in counterfactual temporal alethic-deontic logic. Several (apparently) timeless (or parallel) contrary-to-duty paradoxes are mentioned by Prakken and Sergot (1996).

Here is one example:

Scenario IV: The Dog Warning Sign Scenario (after Prakken and Sergot 1996) Consider the following set of regulations: It ought to be that there is no dog (in a certain place). It ought to be that if there is no dog, there is no warning sign (that warns us about the dog). If there is a dog, it ought to be that there is a warning sign. Suppose, further, that there is a dog. Then all of the following sentences seem to be true:

N-TCTD

TN1   It ought to be that there is no dog.
TN2   It ought to be that if there is no dog, there is no warning sign.
TN3   If there is a dog, it ought to be that there is a warning sign.
TN4   There is a dog.

TN1 expresses a primary obligation and TN3 a contrary-to-duty obligation. The condition in TN3 is satisfied only if the primary obligation expressed by TN1 is violated. In this example, all obligations appear to be timeless or parallel, they appear to be in force simultaneously and the antecedent and consequent in the contrary-to-duty obligation TN3 seem to refer to one and the same time (or perhaps to no particular time at all). Hence, this is an example of a timeless or parallel contrary-to-duty paradox. Let N-TCTD = {TN1, TN2, TN3, TN4}. N-TCTD appears to be consistent, non-redundant, etc. But if this is the case, N-TCTD poses a problem for the counterfactual solution in this paper. For it seems to be impossible to find a plausible formalisation of N-TCTD in counterfactual temporal alethic-deontic logic.

I will now try to show this by considering some possible symbolisations of N-TCTD and some problems with these.

5.2.1 Attempt 1

F1-TCTD

TF1 $O\neg d$
TF2 $O(\neg d \ \Box\rightarrow \neg w)$
TF3 $d \ \Box\rightarrow Ow$
TF4 d

In F1-TCTD, 'd' stands for 'There is a dog' and 'w' for 'There is a warning sign'. All other symbols are interpreted as usual. Let F1-TCTD = {TF1, TF2, TF3, TF4}. This seems to be the most natural interpretation of N-TCTD. But there are at least two problems with this formalisation. First, on this reading, the paradox is reinstated, for F1-TCTD is inconsistent in most plausible counterfactual deontic systems, for example, in **SCTADL**. Second, F1-TCTD generates the pragmatic oddity. In this section, I will prove that this set is inconsistent. In the next, we will see that it also generates the pragmatic oddity.

To prove that a set of sentences {A_1, A_2, ..., A_n} is inconsistent in a tableau system S, we construct an S-tableau that begins with every sentence in this set suffixed in an appropriate way, e.g. A_1, w_0t_0, A_2, w_0t_0, ..., A_n, w_0t_0. If this tableau is closed, that is, if every branch in it is closed, the set is inconsistent in S.

(1) $O\neg d$, w_0t_0
(2) $O(\neg d \ \Box\rightarrow \neg w)$, w_0t_0
(3) $d \ \Box\rightarrow Ow$, w_0t_0
(4) d, w_0t_0
(5) $d \rightarrow Ow$, w_0t_0 [3, DR2]
(6) Ow, w_0t_0 [4, 5, MP]
(7) $sw_0w_1t_0$ [$T - dD$]
(8) $\neg d$, w_1t_0 [1, 7, O]
(9) $\neg d \ \Box\rightarrow \neg w$, w_1t_0 [2, 7, O]
(10) w, w_1t_0 [6, 7, O]
(11) $\neg d \rightarrow \neg w$, w_1t_0 [9, DR2]
(12) $\neg w$, w_1t_0 [8, 11, MP]
(13)   * [10, 12]

So, this symbolisation does not seem to succeed. Let us turn to a second attempt.

5.2.2 Attempt 2

F2-TCTD

TF1 $O\neg d$
TF2b $\neg d \ \Box\rightarrow O\neg w$
TF3 $d \ \Box\rightarrow Ow$
TF4 d

Let F2-TCTD = {TF1, TF2b, TF3, TF4}. From TF3 and TF4, we can deduce Ow, but it is not possible to derive $O\neg w$ from TF1 and TF2b, at least not in most reasonable counterfactual deontic systems. Hence, we cannot derive a contradiction from this set. However, this symbolisation is problematic for at least three different reasons. First, the structure of TF2b does not reflect the structure of the English sentence, TN2. In TN2, the expression 'ought' has wide scope, but in TF2b, the O-operator has narrow scope. Of course, this is not a conclusive argument. But it is nice if our formal sentences mirror our English sentences as close as possible. Similar remarks hold for similar arguments in this section. Second, we cannot derive the ideal obligation that there is no warning sign from TF1 and TF2b. Nevertheless, the proposition that there ought to be no warning sign seems to follow from TN1 and TN2. Third, this solution also generates the pragmatic oddity.

The pragmatic oddity is a problem not only for the second attempt but also for the first attempt above. In every counterfactual deontic system that includes the tableau rule $T - c5$ (see Rönnedal 2016, and the appendix) and, hence, the schema $(A \,\square\!\!\rightarrow B) \rightarrow (A \rightarrow B)$, the sentence $O(\neg d \wedge w)$ is derivable from F2-TCTD (and from F1-TCTD). This is odd since it does not seem to follow that it ought to be that there is no dog *and* that there is a warning sign (that warns us about the dog) from N-TCTD and since it seems that $(A \,\square\!\!\rightarrow B) \rightarrow (A \rightarrow B)$ should hold in every reasonable counterfactual logic. The following semantic tableau shows that $O(\neg d \wedge w)$ is derivable from F2-TCTD (and from F1-TCTD) in the indicated counterfactual deontic systems.

$(1)\ O\neg d,\ w_0 t_0$
$(2)\ \neg d \,\square\!\!\rightarrow O\neg w,\ w_0 t_0$
$(3)\ d \,\square\!\!\rightarrow Ow,\ w_0 t_0$
$(4)\ d,\ w_0 t_0$
$(5)\ \neg O(\neg d \wedge w),\ w_0 t_0$
$(6)\ P\neg(\neg d \wedge w),\ w_0 t_0\ [5, \neg O]$
$(7)\ sw_0 w_1 t_0\ [6, P]$
$(8)\ \neg(\neg d \wedge w),\ w_1 t_0\ [6, P]$
$(9)\ \neg d,\ w_1 t_0\ [1, 7, O]$
$(10)\ d \rightarrow Ow,\ w_0 t_0\ [3, DR2]$
$(11)\ Ow,\ w_0 t_0\ [4, 10, MP]$
$(12)\ w,\ w_1 t_0\ [11, 7, O]$

$\swarrow$ $\searrow$

$(13)\ \neg\neg d,\ w_1 t_0\ [8, \neg\wedge]$ $(14)\ \neg w,\ w_1 t_0\ [8, \neg\wedge]$
$(15)\ *\ [9, 13]$ $(16)\ *\ [12, 14]$

In conclusion, this attempt does not seem to succeed either. Our third attempt below avoids the pragmatic oddity. So, maybe this symbolisation can be used to solve timeless or parallel contrary-to-duty paradoxes.

5.2.3 Attempt 3

F3-TCTD

TF1 $O\neg d$
TF2 $O(\neg d \,\square\!\!\rightarrow \neg w)$

TF3b $O(d \,\square\!\rightarrow w)$
TF4 d

Let F3-TCTD = {TF1, TF2, TF3b, TF4}. In this set, not only TN2 but also TN3 is represented by a sentence where the O-operator has wide scope. From this set, we can derive $O\neg w$ from TF1 and TF2, but Ow does not follow from TF3b and TF4. The set is not inconsistent. Moreover, the solution does not generate the pragmatic oddity. $O(\neg d \wedge w)$ does not follow from F3-TCTD. Nevertheless, there are at least three problems with this symbolisation. First, $O(d \,\square\!\rightarrow w)$ does not reflect the structure of the natural language sentence, TN3. In TN3, the word 'ought' has narrow scope, but the O-operator in $O(d \,\square\!\rightarrow w)$ has wide scope. Second, we cannot derive the 'actual' obligation that there ought to be a warning sign from TF3b and TF4. But it seems to follow that there ought to be a warning sign from TN3 and TN4. Third, $O(d \,\square\!\rightarrow w)$ does not seem to be true. This is problematic since we have stipulated that TN3 is true in our scenario. So, the symbolisation of this sentence should also be true. We will now consider the argument for this claim.

All of the following sentences seem to be true: $O(\neg d \,\square\!\rightarrow \neg w)$, $\neg d \,\square\!\rightarrow O\neg w$ and $d \,\square\!\rightarrow Ow$. But $O(d \,\square\!\rightarrow w)$ appears to be false. According to the standard truth-conditions for counterfactuals, $A \,\square\!\rightarrow B$ is true in a possible world ω if and only if B is true in every possible world that is as close as (as similar as) possible to ω in which A is true; and OA is true in a possible world ω if and only if A is true in every possible world that is deontically accessible from ω. If we think of the truth-conditions in this way, $O(d \,\square\!\rightarrow w)$ is true in $\omega_@$ (our world) just in case $d \,\square\!\rightarrow w$ is true in all ideal worlds (i.e. in all possible worlds that are deontically accessible from $\omega_@$), that is, if and only if, in every ideal world, ω' deontically accessible from $\omega_@$, w is true in all the worlds that are as close to ω' as possible in which d is true. But in all ideal worlds, there is no dog (in the specified place), and in all ideal worlds, if there is no dog, there is no warning sign (warning us about the dog). From this, it follows that in all ideal worlds, there is no warning sign. Accordingly, in all ideal worlds, there is no dog and there is no warning sign. Take an ideal world, say ω'. In the closest d world(s) to ω', $\neg w$ seems to be true (since $\neg w$ is true in ω'). If this is correct, d and $\neg w$ are true in one of the closest d worlds to ω'. So, $d \,\square\!\rightarrow w$ is not true in ω'. Hence, $O(d \,\square\!\rightarrow w)$ is not true in $\omega_@$ (in our world). In conclusion, if this argument is sound, we cannot avoid the contrary-to-duty paradox by using the symbolisation, F3-TCTD.

5.2.4 Attempt 4

I will mention one more attempt to symbolise N-TCTD. So far, we have not used any temporal operators in our formalisations. This reflects the idea that we are dealing with a *timeless* contrary-to-duty paradox. Might it perhaps be possible to solve the paradox if we introduce some temporal operator(s) in the symbolisation? The antecedent and the consequent in the two conditional norms expressed by TN2 and TN3 are supposed to 'refer' to the same moment in time. So, the only plausible formalisation that includes some temporal operators seems to be the following:

F4-TCTD

TCF1 $Rt_1ORt_1\neg d$
TCF2 $Rt_1O(Rt_1\neg d \,\square\!\rightarrow Rt_1\neg w)$

TCF3 $Rt_1(Rt_1d \,\square\!\rightarrow Rt_1ORt_1w)$
TCF4 Rt_1d

Where 't_1' refers to some arbitrary point in time. Unfortunately, this does not help. For F4-TCTD = {TCF1, TCF2, TCF3, TCF4} is also inconsistent. From TCF1 and TCF2, we can derive the sentence $Rt_1ORt_1\neg w$, and TCF3 and TCF4 entail the formula Rt_1ORt_1w. Together these imply $Rt_1ORt_1\neg w \;\wedge\; Rt_1ORt_1w$, which is a kind of moral dilemma. Moreover, F4-TCTD entails $Rt_1ORt_1(\neg d \wedge w)$, which is a version of the pragmatic oddity. So, this strategy does not work either.

Therefore, the only reasonable conclusion seems to be that the counterfactual solution discussed in this paper cannot be used to solve timeless or parallel contrary-to-duty paradoxes. Nonetheless, maybe the arguments discussed so far are not decisive. I will now consider one possible counterargument.

5.2.5 Possible Reply

So far we have assumed that N-TCTD is consistent, and that it does not generate any (moral) dilemma. But this is not obviously correct. Is it really true that this set does not entail any contradiction or any (moral) dilemma? One way to respond to the arguments above is to deny this. It is possible that N-TCTD is inconsistent and/or entails a (moral) dilemma. And if this is the case, it is not particularly surprising that our symbolisation, F1-TCTD, which is the most natural rendering of N-TCTD, is inconsistent. In fact, if this is the case, our formalisation of N-TCTD *should* be inconsistent. The intuition that N-TCTD does not entail any contradiction can be explained by the fact that N-TCTD is similar to afterward contrary-to-duty paradoxes, such as N-CTD, which are not inconsistent. This fools us into believing that N-TCTD is also consistent. But, in fact, it is not. Consequently, it is not necessarily the case that the timeless or parallel contrary-to-duty paradoxes refute the counterfactual solution presented in this paper. At the very least, we should not dismiss this possibility without very good arguments. The fact that a certain timeless contrary-to-duty scenario *appears* to be consistent is not enough to conclude that it *is* consistent.[5]

Hence, let us turn to our third argument, which seems to be the most problematic for the counterfactual solution.

---

[5] This reply can be used to meet the following possible counterargument to the topic in this paper. Suppose that pure counterfactual deontic logic can handle the contrary-to-duty paradoxes satisfactorily. Then there is no need to extend it with time, as far as the paradoxes are concerned. If the pure counterfactual deontic logic cannot handle the paradoxes, then its tense version will not either. Whatever they are, all the problems the first run into will arise for parallel scenarios, where the consequent of the 'obligation operator' and its antecedent pertain to the same point in time. In either case, the basic framework is disqualified. Moreover, if this is true, there is no need to also consider beforehand scenarios. Nevertheless, it is quite possible that afterward, and beforehand contrary-to-duty scenarios are consistent, while timeless scenarios are not. If this is true, and I am inclined to believe that it is (perhaps with some exceptions), it does not matter if the symbolisation of N-TCTD is inconsistent. It is, however, problematic if the counterfactual solution cannot handle beforehand contrary-to-duty paradoxes. Even if parallel contrary-to-duty scenarios were consistent, we would want to know if the counterfactual solution can handle beforehand paradoxes.

5.3 Argument 3: The Counterfactual Solution cannot Solve Beforehand Contrary-to-Duty Paradoxes

According to the third argument against the counterfactual solution, we cannot use counterfactual temporal alethic-deontic logic to solve beforehand contrary-to-duty paradoxes. This seems to be the most difficult problem for this kind of solution. Consider the following scenario:

Scenario V: The (Beforehand) Promise (Contrary-to-Duty) Paradox Scenario V is similar to Scenario I. Having said that, Scenario I involves an *afterward* contrary-to-duty obligation while Scenario V is concerned with a *beforehand* contrary-to-duty obligation. On Monday, you promise a friend to meet her on Friday to help her with some task. Hence, you ought to keep your promise. If you will not keep your promise, you ought to call your friend on Wednesday, tell her that you will not come and apologise. In this example, the following sentences all seem to be true:

N-BCTD

BN1   (On Monday it is true that) You ought to keep your promise (and see your friend on Friday).

BN2   (On Monday it is true that) It ought to be that if you keep your promise, you do not call your friend (on Wednesday).

BN3   (On Monday it is true that) If you do not keep your promise (i.e. if you do not see your friend on Friday and help her out), you ought to call her (on Wednesday).

BN4   (On Monday it is true that) You do not keep your promise (on Friday), that is, you will not see your friend.

Let N-BCTD = {BN1, BN2, BN3, BN4}. BN3 is a contrary-to-duty obligation (or expresses a contrary-to-duty obligation). If the condition is true, the primary obligation that you should keep your promise (expressed by BN1) is not fulfilled. Moreover, it is a beforehand contrary-to-duty obligation. If you will not be keeping your promise on Friday, that is, if you are not going to see your friend and help her, you ought to call her beforehand, on Wednesday. Then she will be less frustrated and she might perhaps find someone else that can help her. N-BCTD seems to be consistent; it does not seem possible to deduce any contradiction from this set. However, it is difficult to find any reasonable symbolisation of N-BCTD. This is an example of a beforehand contrary-to-duty paradox. I will now consider some attempts to solve this paradox in counterfactual temporal alethic-deontic logic. All of these attempts are problematic for various reasons.

5.3.1 Attempt 1

F1-BCTD

FB1 Rt_1ORt_3k
FB2 $Rt_1O(Rt_3k \mathbin{\square\!\rightarrow} Rt_2\neg c)$
FB3 $Rt_1(Rt_3\neg k \mathbin{\square\!\rightarrow} Rt_1ORt_2c)$
FB4 $Rt_1Rt_3\neg k$

All symbols are interpreted as above, and 'c' says: 'You call your friend (tell her that you will not come and apologise)'. Let F1-BCTD = {FB1, FB2, FB3, FB4}. Is this attempt successful? It seems that it is not for at least two reasons. First, F1-BCTD is inconsistent in every system that includes the non-basic tableau rules $T - c5$ and $T - dD$. ($T - dD$ rules out moral dilemmas.) Second, we can derive the following version of the pragmatic oddity from F1-BCTD: $Rt_1O(Rt_3k \wedge Rt_2c)$, which says that it is true on Monday that you ought to keep your promise on Friday *and* call your friend on Wednesday and tell her that you are not coming on Friday. This is clearly a counterintuitive conclusion. The following semantic tableau establishes that F1-BCTD is inconsistent:

(1) Rt_1ORt_3k, w_0t_0
(2) $Rt_1O(Rt_3k \,\square\!\rightarrow Rt_2\neg c)$, w_0t_0
(3) $Rt_1(Rt_3\neg k \,\square\!\rightarrow Rt_1ORt_2c)$, w_0t_0
(4) $Rt_1Rt_3\neg k$, w_0t_0
(5) $Rt_3\neg k \,\square\!\rightarrow Rt_1ORt_2c$, w_0t_1 [3, Rt]
(6) $Rt_3\neg k \rightarrow Rt_1ORt_2c$, w_0t_1 [5, $DR2$]
(7) $Rt_3\neg k$, w_0t_1 [4, Rt]
(8) Rt_1ORt_2c, w_0t_1 [6, 7, MP]
(9) ORt_3k, w_0t_1 [1, Rt]
(10) $O(Rt_3k \,\square\!\rightarrow Rt_2\neg c)$, w_0t_1 [2, Rt]
(11) ORt_2c, w_0t_1 [8, Rt]
(12) $sw_0w_1t_1$ [$T - dD$]
(13) Rt_3k, w_1t_1 [9, 12, O]
(14) $Rt_3k \,\square\!\rightarrow Rt_2\neg c$, w_1t_1 [10, 12, O]
(15) $Rt_3k \rightarrow Rt_2\neg c$, w_1t_1 [14, $DR2$]
(16) Rt_2c, w_1t_1 [11, 12, O]
(17) $Rt_2\neg c$, w_1t_1 [13, 15, MP]
(18) $\neg c$, w_1t_2 [17, Rt]
(19) c, w_1t_2 [16, Rt]
(20)   * [18, 19]

Note that it does not help to replace FB3, $Rt_1(Rt_3\neg k \,\square\!\rightarrow Rt_1ORt_2c)$, by $Rt_1(Rt_3\neg k \,\square\!\rightarrow ORt_2c)$ in the derivation above. We can still prove that the resulting set is inconsistent.

The pragmatic oddity is also a problem for our second attempt to solve the beforehand contrary-to-duty paradox. The proof is the same in both cases (see below). So, let us turn to this attempt.

5.3.2 Attempt 2

In the symbolisation below, we have replaced FB2 [$Rt_1O(Rt_3k \,\square\!\rightarrow Rt_2\neg c)$] by FB2b [$Rt_1(Rt_3k \,\square\!\rightarrow Rt_1ORt_2\neg c)$]. According to FB2, the O-operator has wide scope, but according to FB2b, it has narrow scope.

F2-BCTD

FB1 Rt_1ORt_3k
FB2b $Rt_1(Rt_3k \,\square\!\rightarrow Rt_1ORt_2\neg c)$

FB3 $Rt_1(Rt_3\neg k \;\square\!\!\rightarrow Rt_1ORt_2c)$
FB4 $Rt_1Rt_3\neg k$

Let F2-BCTD = {FB1, FB2b, FB3, FB4}. Then we cannot derive a contradiction in the way we did above. So, in this sense, F2-BCTD solves the paradox. Still, there are at least two problems with this solution. Firstly, FB2b does not reflect the structure of the English sentence, BN2. In BN2, the word 'ought' has wide scope, but the O-operator in FB2b has narrow scope. Secondly, and more problematically, F2-BCTD entails the following form of the pragmatic oddity, $Rt_1O(Rt_3k \wedge Rt_2c)$. Recall that this formula says that it is true on Monday that you ought to keep your promise on Friday *and* call your friend on Wednesday and tell her that you are not coming on Friday. But this proposition does not seem to follow from N-BCTD. As I mentioned above, this is also a problem for our first attempt to solve the paradox. The following semantic tableau shows that $Rt_1O(Rt_3k \wedge Rt_2c)$ is derivable from F2-BCTD (and from F1-BCTD):

(1) Rt_1ORt_3k, w_0t_0
(2) $Rt_1(Rt_3\neg k \;\square\!\!\rightarrow Rt_1ORt_2c)$, w_0t_0
(3) $Rt_1Rt_3\neg k$, w_0t_0
(4) $\neg Rt_1O(Rt_3k \wedge Rt_2c)$, w_0t_0
(5) $Rt_3\neg k \;\square\!\!\rightarrow Rt_1ORt_2c$, w_0t_1 [2, Rt]
(6) $Rt_3\neg k \rightarrow Rt_1ORt_2c$, w_0t_1 [5, DR2]
(7) $Rt_3\neg k$, w_0t_1 [3, Rt]
(8) Rt_1ORt_2c, w_0t_1 [6, 7, MP]
(9) $\neg O(Rt_3k \wedge Rt_2c)$, w_0t_1 [4, DR1]
(10) $P\neg(Rt_3k \wedge Rt_2c)$, w_0t_1 [9, $\neg O$]
(11) $sw_0w_1t_1$ [10, P]
(12) $\neg(Rt_3k \wedge Rt_2c)$, w_1t_1 [10, P]

↙ ↘

(13) $\neg Rt_3k$, w_1t_1 [12, $\neg\wedge$]   (14) $\neg Rt_2c$, w_1t_1 [12, $\neg\wedge$]
(15) $\neg k$, w_1t_3 [13, DR1]   (16) $\neg c$, w_1t_2 [14, DR1]
(17) ORt_3k, w_0t_1 [1, Rt]   (18) ORt_2c, w_0t_1 [8, Rt]
(19) Rt_3k, w_1t_1 [17, 11, O]   (20) Rt_2c, w_1t_1 [18, 11, O]
(21) k, w_1t_3 [19, Rt]   (22) c, w_1t_2 [20, Rt]
(23) * [15, 21]   (24) * [16, 22]

Note that it does not help to replace FB3, $Rt_1(Rt_3\neg k \;\square\!\!\rightarrow Rt_1ORt_2c)$, by $Rt_1(Rt_3\neg k \;\square\!\!\rightarrow ORt_2c)$ in the proof above. We can still derive the pragmatic oddity with this sentence instead of FB3.

5.3.3 Attempt 3

Attempt 3 is similar to Attempt 1. Although, we have replaced FB3 [$Rt_1(Rt_3\neg k \;\square\!\!\rightarrow Rt_1ORt_2c)$] by FB3b [$Rt_1O(Rt_3\neg k \;\square\!\!\rightarrow Rt_2c)$]. In FB3b, the O-operator has wide scope instead of narrow scope as in FB3.

F3-BCTD

FB1 Rt_1ORt_3k
FB2 $Rt_1O(Rt_3k \;\square\!\rightarrow Rt_2\neg c)$
FB3b $Rt_1O(Rt_3\neg k \;\square\!\rightarrow Rt_2c)$
FB4 $Rt_1Rt_3\neg k$

Let F3-BCTD = {FB1, FB2, FB3b, FB4}. This seems to be one of the most plausible symbolisations of N-BCTD in counterfactual temporal alethic-deontic logic. The set is consistent and non-redundant, and it has several other attractive features too (in most reasonable systems). Nevertheless, there are at least four problems with this attempt. First, FB3b does not reflect the structure of BN3. In the English sentence, 'ought' has narrow scope; in FB3b, the deontic operator O has wide scope. Second, we cannot derive any 'actual' obligation from this set, that is, we cannot derive the proposition that you ought to call your friend from FB3b and FB4. But intuitively, this claim seems to follow from BN3 and BN4. A possible reply is to replace FB4 by FB4b ($Rt_1\square Rt_3\neg k$). In other words, we assume that it is true at t_1 that it is settled at t_3 that $\neg k$. Then we can derive the sentence, Rt_1ORt_2c, from FB3b and FB4b. However, the set F3b-BCTD = {FB1, FB2, FB3b, FB4b} is inconsistent. We can derive both Rt_1ORt_2c and $Rt_1ORt_2\neg c$ from F3b-BCTD and, from this, it follows that $Rt_1(ORt_2c \wedge ORt_2\neg c)$ (and that $Rt_1O(Rt_2c \wedge Rt_2\neg c)$), which is inconsistent with the idea that there are no moral dilemmas. (It is also possible to use semantic tableau directly to prove that F3b-BCTD is inconsistent. Though, we will not labour the details.) So, this move does not appear to succeed. Third, in our solution to the afterward contrary-to-duty paradox, we used the sentence $Rt_1(Rt_2\neg k \;\square\!\rightarrow Rt_2ORt_3a)$. In this formula, the O-operator has narrow scope. It seems a little bit odd that we should use a sentence with a wide scope O-operator to symbolise beforehand contrary-to-duty obligations and a sentence with a narrow scope O-operator to symbolise an afterward contrary-to-duty obligation.

Fourth, $Rt_1O(Rt_3\neg k \;\square\!\rightarrow Rt_2c)$ does not seem to be true. The most natural model of Scenario V appears to be the following. We have four possible worlds $\omega_1 - \omega_4$ and three moments of time $\tau_1 - \tau_3$. At τ_1, every possible world is alethically accessible from every possible world. At τ_2, ω_1 is alethically accessible from ω_2 and ω_2 from ω_1, and ω_3 is alethically accessible from ω_4 and ω_4 from ω_3. Every possible world is alethically accessible to itself at every moment of time. At τ_1, ω_3 is deontically accessible from every possible world. At τ_2, ω_3 is deontically accessible from ω_3 and ω_4, and ω_1 is deontically accessible from ω_1 and ω_2. At τ_3, every possible world is deontically accessible to itself. c is true in ω_1 and ω_2 and false in ω_3 and ω_4 at τ_2. k is true in ω_2 and ω_3 and false in ω_1 and ω_4 at τ_3. $v(t_1) = \tau_1$, $v(t_2) = \tau_2$ and $v(t_3) = \tau_3$. Furthermore, ω_4 is more similar to ω_3 than is ω_1. In this model, every sentence in F3-BCTD is true (in ω_1 and ω_4 at τ_1) except FB3b. But intuitively, BN3 is true. So, we want our symbolisation of this sentence to be true also. Moreover, it seems difficult to find any other plausible model of Scenario V. This is a problem for Attempt 3.

5.3.4 Attempt 4

In F1-BCTD and F2-BCTD, the consequent in the contrary-to-duty obligation begins with Rt_1. That is, we assume that the obligation to call your friend given that you will not keep your promise is in force already at t_1, that is, on Monday. What happens if we

replace FB3 with FB3c: $Rt_1(Rt_3\neg k \,\square\rightarrow Rt_2ORt_2c)$, which says that the obligation to call your friend is in force at t_2 given that you will not keep your promise? Does this solve the beforehand contrary-to-duty paradox? Consider the following symbolisation of N-BCTD:

F4-BCTD

FB1 Rt_1ORt_3k
FB2 $Rt_1O(Rt_3k \,\square\rightarrow Rt_2\neg c)$
FB3c $Rt_1(Rt_3\neg k \,\square\rightarrow Rt_2ORt_2c)$
FB4 $Rt_1Rt_3\neg k$

Let F4-BCTD = {FB1, FB2, FB3c, FB4}. $Rt_1ORt_2\neg c$ follows from FB1 and FB2, and from FB4 and FB3c we can derive Rt_2ORt_2c. But $Rt_1ORt_2\neg c$ is not inconsistent with Rt_2ORt_2c. $Rt_1ORt_2\neg c$ says that it is true at time t_1 that it is obligatory that you do not call your friend at time t_2; and Rt_2ORt_2c says that it is true at time t_2 that it is obligatory that you call your friend at t_2, that is, on Wednesday. Obligations may change over time and something may be obligatory at one time even though it is not obligatory at a later time. F4-BCTD also has some other attractive features; it is, for example, non-redundant.

Nevertheless, Attempt 4 is problematic for at least one reason. In FB3c, the consequent Rt_2ORt_2c is not an obligation that is 'future oriented' as genuine norms are or should be according to many deontic logicians. An obligation is future oriented if the content of the obligation involves a time that is later than the obligation itself. The consequent Rt_2ORt_3a in $Rt_1(Rt_2\neg k \,\square\rightarrow Rt_2ORt_3a)$, for instance, is future oriented since t_3 is later than t_2. Here are some other examples of future-oriented obligations: OFk, OGk, where F ('It will at some time in the future be the case that') and G ('It is always going to be the case that') are temporal operators that are future oriented.

In any counterfactual temporal alethic-deontic system that includes the rules $T - dD$, $T - MO$, $T - FT$ and $T - BT$, the sentence Rt_2ORt_2c is equivalent to Rt_2c. Hence, the proposition that it is true at t_2 that it ought to be the case that you call your friend at t_2 is equivalent to the proposition that it is true at t_2 that you call your friend. Consequently, $Rt_1(Rt_3\neg k \,\square\rightarrow Rt_2ORt_2c)$ (FB3c) is equivalent to $Rt_1(Rt_3\neg k \,\square\rightarrow Rt_2c)$. This is a strange conclusion. The following semantic tableau proves that Rt_2ORt_2c is equivalent to Rt_2c in the indicated systems. ('$\leftrightarrow$' stands for material equivalence.)

$$(1)\ \neg(Rt_2ORt_2c \leftrightarrow Rt_2c),\ w_0t_0$$

<div style="display:flex; justify-content:space-between;">

(2) Rt_2ORt_2c, w_0t_0 [1, $\neg\leftrightarrow$]
(4) $\neg Rt_2c$, w_0t_0 [1, $\neg\leftrightarrow$]
(6) $\neg c$, w_0t_2 [4, $DR1$]
(8) ORt_2c, w_0t_2 [2, Rt]
(10) $sw_0w_1t_2$ [$T-dD$]
(12) Rt_2c, w_1t_2 [8, 10, O]
(14) c, w_1t_2 [12, Rt]
(16) $rw_0w_1t_2$ [10, $T-MO$]
(18) c, w_0t_2 [14, 16, $T-BT$]
(20) * [6, 18]

(3) $\neg Rt_2ORt_2c$, w_0t_0 [1, $\neg\leftrightarrow$]
(5) Rt_2c, w_0t_0 [1, $\neg\leftrightarrow$]
(7) c, w_0t_2 [5, Rt]
(9) $\neg ORt_2c$, w_0t_2 [3, $DR1$]
(11) $P\neg Rt_2c$, w_0t_2 [9, $\neg O$]
(13) $sw_0w_1t_2$ [11, P]
(15) $\neg Rt_2c$, w_1t_2 [11, P]
(17) $\neg c$, w_1t_2 [15, $DR1$]
(19) $rw_0w_1t_2$ [13, $T-MO$]
(21) c, w_1t_2 [7, 19, $T-FT$]
(22) * [17, 21]

</div>

This result is a problem for Attempt 4 if it is reasonable to assume that our counterfactual temporal alethic-deontic tableau system *should* include the rules that are used in the derivation. And for many applications, these rules seem plausible, for example, if we consider the past and the present as settled or historically necessary, as many deontic logicians do.

We have now considered four attempts to solve the beforehand contrary-to-duty paradox. We have seen that all of these solutions are problematic. It appears to be difficult to find any other formalisation of N-BCTD in counterfactual temporal alethic-deontic logic that is plausible. This is a serious problem for the counterfactual solution. In Section 5.2.5, I suggested that a possible reply to the argument from parallel contrary-to-duty paradoxes against the counterfactual solution is to reject the idea that N-TCTD is consistent. A similar response to the problem with beforehand contrary-to-duty paradoxes does not seem to be plausible. The problem with N-TCTD is that all obligations are supposed to be in force at the same time. But in beforehand contrary-to-duty paradoxes, this is not the case. It is intuitively more plausible to think that N-BCTD is consistent than to think that N-TCTD is consistent. Therefore, Argument 3 seems to be more problematic than Argument 2.

One might argue that it is enough that the approach used in this paper can solve afterward contrary-to-duty paradoxes. Perhaps one and the same solution cannot take care of all different versions of this puzzle. The fact that the counterfactual solution has problems with parallel and beforehand paradoxes is, therefore, not a decisive argument against it. But if there is some other approach that can be used to solve not only afterward contrary-to-duty paradoxes but also beforehand contrary-to-duty paradoxes, such a solution seems preferable. And, in fact, there might be such a solution. Several philosophers have suggested that contrary-to-duty paradoxes can be solved in a system that combines dyadic deontic logic and some type of temporal logic (e.g. van Eck 1982; Loewer and Belzer 1983; Feldman 1986, 1990; Åqvist 2003). Moreover, it seems to be the case that a solution of this kind can take care of afterward *and* beforehand contrary-to-duty paradoxes. This suggests that a solution of this type might be, all things considered, more plausible than the counterfactual solution, even though it too has problems with parallel contrary-to-duty paradoxes.

6 Conclusion

In this paper, I have discussed some examples of the so-called contrary-to-duty (obligation) paradox, a puzzle that has been investigated by deontic logicians for more than fifty years. A contrary-to-duty obligation is an obligation telling us what ought to be the case if something forbidden is true. Here are some examples of sentences that can be used to express such obligations: 'If he did it, he should be ashamed of himself', 'If the resources are unfairly distributed, they should be redistributed', 'If you are going to take drugs, you ought to use clean needles'. In each case, the condition is supposed to violate a primary obligation. He should not have done it, but if he did, he should be ashamed of himself. The resources should not be unfairly distributed, but if they are, they should be redistributed. You should not take drugs; but if you are going to take drugs, you should use clean needles. Contrary-to-duty obligations are important in our moral and legal thinking. Therefore, we want to be able to find an adequate

symbolisation of such obligations in some logical system, a task that has turned out to be difficult. This is clearly illustrated by the so-called contrary-to-duty (obligation) paradox. I have investigated and evaluated one kind of solution to this problem that has been suggested in the literature, a solution that I called the 'counterfactual solution'. I used some recent systems introduced by Rönnedal (2016) that combine not only counterfactual logic and deontic logic, but also temporal logic in my analysis of the paradox. I have argued that the counterfactual solution has many attractive features and that it can give a fairly satisfactory answer to some examples of the contrary-to-duty paradox, but that it nevertheless has some serious problems. The conclusion is that, notwithstanding the many attractive features of the counterfactual solution, there seem to be other approaches to the paradox that are more promising, especially a kind of solution that uses a combination of dyadic deontic logic and temporal logic.

Appendix

This appendix includes a description of the formal systems that are used in this paper. Only the most essential information is included; for more information, see Rönnedal (2016). I use the same names of rules, conditions, etc. as in Rönnedal (2016), but I sometimes use slightly different symbols. Not every rule that is included in our systems is described in detail; some rules that are not explicitly used in the paper are omitted.

Syntax

Alphabet

(i) A set Prop of propositional letters p, q, r, s, p_1, q_1, r_1, s_1, p_2, q_2, r_2, s_2, ..., (ii) A set NT of names of times t_0, t_1, t_2, t_3 ..., (iii) primitive truth-functional connectives ¬ (negation), ∧ (conjunction), ∨ (disjunction), → (material implication) and ↔ (material equivalence), (iv) alethic operators U, M, □, ◇, ⊡ and ⟡, (v) temporal operators $\underline{R}$, $\underline{A}$, $\underline{S}$, $\underline{G}$, $\underline{H}$, $\underline{F}$ and $\underline{P}$, (vi) deontic operators O and P, (vii) counterfactual operators □→ and ◇→, and (viii) ⊤ (verum), ⊥ (falsum) and the brackets), (.

Language

The language L is the set of well-formed formulas (wffs) generated by the usual clauses for propositional letters and propositionally compound sentences, and the following clauses: (i) if A is a wff, then UA ('it is universally (or absolutely) necessary that A'), MA ('it is universally (or absolutely) possible that A'), □A ('it is historically necessary (or settled) that A'), ◇A ('it is historically possible that A'), ⊡ A ('it is temporally (synchronistically) necessary that A'), ⟡A ('it is temporally (synchronistically) possible that A'), $\underline{A}A$ ('it is always the case that A'), $\underline{S}A$ ('it is some time the case that A'), $\underline{G}A$ ('it is always going to be

the case that A'), $\underline{H}A$ ('it has always been the case that A'), $\underline{F}A$ ('it will some time in the future be the case that A'), $\underline{P}A$ ('it was some time in the past the case that A'), OA ('it ought to be the case that A') and $\overline{P}A$ ('it is permitted that A') are wffs, (ii) if A is a wff and t is in NT, then RtA ('it is realised at time t that A') is a wff, (iii) if A and B are wffs, so are ($A \,\square\!\rightarrow B$) ('If it were the case that A, then it would be the case that B') and ($A \,\diamond\!\rightarrow B$) ('If it were the case that A, then it might be the case that B'), and (iv) nothing else is a wff.

Semantics

Counterfactual temporal alethic-deontic frames and models

A frame F is a relational structure $\langle W, T, <, R, S, \{R_A: A \in L\}, \geq\rangle$, where W is a non-empty set of possible worlds, T is a non-empty set of times, $<$ is a binary relation on T ($< \subseteq T \times T$), R and S are two ternary (alethic or deontic) accessibility relations ($R \subseteq W \times W \times T$ and $S \subseteq W \times W \times T$), $\{R_A: A \in L\}$ is a set of ternary counterfactual accessibility relations, one for each sentence, A, in L ($R_A \subseteq W \times W \times T$), and $\geq$ is a ternary similarity relation defined over the elements in W ($\geq \subseteq W \times W \times W$).

Informally, $\tau < \tau'$ says that the time τ is before the time τ', $R\omega\omega'\tau$ says that the possible world ω' is alethically accessible from the possible world ω at time τ, $S\omega\omega'\tau$ says that ω' is deontically accessible from ω at τ, $R_A\omega\omega'\tau$ says that ω' is A-accessible from ω at τ, and $\omega \geq_{\omega'} \omega''$ says that ω is at least as similar to ('near' to) ω' as is ω''. To 'decide' whether $\omega \geq_{\omega'} \omega''$, we place ourselves in ω', and ask which world is more similar to ours, ω or ω''. We can think of $\omega \geq_{\omega'} \omega''$ as a dyadic similarity relation relativised to ω'.

A model M is a triple $\langle F, V, v\rangle$ where: (i) F is a frame; (ii) V is a valuation or interpretation function, which to every propositional letter p in Prop assigns a subset of $W \times T$, i.e. a set of ordered pairs $\langle\omega, \tau\rangle$, where $\omega \in W$ and $\tau \in T$; and (iii) v is a function which to each temporal name in NT assigns a time in T.

Let $M, \omega, \tau \Vdash A$ abbreviate 'A is true in the possible world ω at the time τ in the model M' (or 'A is true at the pair $\langle\omega, \tau\rangle$ in M'), and let x, y, z, w, etc. range over possible worlds in W, and t over times in T. The counterfactual accessibility relations can be defined in terms of the similarity relation over our possible worlds in the following way:

$$\gamma 0.\text{For every } A, R_A xyt \text{ iff } M, y, t \Vdash A \text{ and} \forall z \left(M, z, t \Vdash A \rightarrow y \geq_x z\right).$$

Intuitively, this condition says that the possible world y is A-accessible from the possible world x at the time t iff y is one of the closest worlds to x in which A is true at t.

Truth in a model, validity, satisfiability, logical consequence, etc.

Here are the truth-conditions of the various sentences in our language (the omitted conditions are straightforward):

> (i) $M, \omega, \tau \Vdash p$ iff $\langle\omega, \tau\rangle \in V(p)$ for any p in Prop.
> (ii) $M, \omega, \tau \Vdash A \wedge B$ iff $M, \omega, \tau \Vdash A$ and $M, \omega, \tau \Vdash B$. Other truth-functional connectives: similar.

(iii) $M, \omega, \tau \Vdash \Box A$ iff for all $\omega' \in W$ s.t. $R\omega\omega'\tau$: $M, \omega', \tau \Vdash A$. The conditions for (iv) $\Diamond$, (v) $\boxdot$, and (vi) $\diamondsuit$ are similar.

(vii) $M, \omega, \tau \Vdash \underline{A}A$ iff for all $\tau' \in T$: $M, \omega, \tau' \Vdash A$. (viii) $\underline{S}$, similar.
(ix) $M, \omega, \tau \Vdash \underline{G}A$ iff for all $\tau' \in T$ s.t. $\tau < \tau'$: $M, \omega, \tau' \Vdash A$. The conditions for (x) $\underline{F}$, (xi) $\underline{H}$, and (xii) $\underline{P}$ are similar.
(xiii) $M, \omega, \tau \Vdash Rt'A$ iff $M, \omega, v(t') \Vdash A$, for all $t' \in NT$.
(xiv) $M, \omega, \tau \Vdash OA$ iff for all $\omega' \in W$ s.t. $S\omega\omega'\tau$: $M, \omega', \tau \Vdash A$.
(xv) $M, \omega, \tau \Vdash PA$ iff for some $\omega' \in W$ s.t. $S\omega\omega'\tau$: $M, \omega', \tau \Vdash A$.
(xvi) $M, \omega, \tau \Vdash UA$ iff for all $\omega' \in W$ and for all $\tau' \in T$: $M, \omega', \tau' \Vdash A$. (xvii) M, similar.
(xviii) $M, \omega, \tau \Vdash A \Box\!\rightarrow B$ iff for all $\omega' \in W$ s.t. $R_A\omega\omega'\tau$: $M, \omega', \tau \Vdash B$.
(xix) $M, \omega, \tau \Vdash A \Diamond\!\rightarrow B$ iff for some $\omega' \in W$ s.t. $R_A\omega\omega'\tau$: $M, \omega', \tau \Vdash B$.

$A \Box\!\rightarrow B$ is true in a world ω at time τ iff B is true in all possible worlds that are A-accessible from ω at τ. $A \Diamond\!\rightarrow B$ is true in a world ω at time τ iff B is true in at least one possible world that is A-accessible from ω at τ. Informally, a would-counterfactual is often said to be true (roughly) just in case the consequent is true in the nearest possible world(s) where the antecedent is true. If we define the counterfactual accessibility relations in terms of the similarity relation as in $\gamma 0$, we can extend this basic idea to our temporal alethic-deontic systems. Then the following truth-conditions are derivable:

$M, \omega, \tau \Vdash A \Box\!\rightarrow B$ iff for all $\omega' \in W$ s.t. $M, \omega', \tau \Vdash A$ and for all $\omega''(M, \omega'', \tau \Vdash A \rightarrow \omega' \geq_\omega \omega'')$: $M, \omega', \tau \Vdash B$.

$M, \omega, \tau \Vdash A \Diamond\!\rightarrow B$ iff for some $\omega' \in W$ s.t. $M, \omega', \tau \Vdash A$ and for all $\omega''(M, \omega'', \tau \Vdash A \rightarrow \omega' \geq_\omega \omega'')$: $M, \omega', \tau \Vdash B$.

Intuitively, this means that $A \Box\!\rightarrow B$ is true in the world ω at the time τ iff in every world ω' that is as close to ω as possible in which A is true at τ, B is true at τ. $A \Diamond\!\rightarrow B$ is true in the world ω at the time τ iff there is a world ω' that is as close to ω as possible in which A is true at τ, in which B is true at τ. Other basic semantic concepts like validity, satisfiability, logical consequence etc. are defined as usual.

Conditions on models, classes of models and the logic of a class of models

By imposing various conditions on the accessibility relations and the valuation function in a model it is possible to obtain a categorisation of the set of all models into various kinds. These classes of models can be used to define a large set of logical system. The set of all sentences (in the language L) that are valid in a class of models M is called the logical system of M.

R may, for example, be reflexive, symmetric, transitive and/or Euclidean, S may be serial, transitive, Euclidean, secondarily reflexive and/or secondarily symmetric, and the temporal accessibility relation $<$ may, for example, be transitive, comparable and/or non-branching towards the future or the past. In addition, we may impose various conditions on $\{R_A: A \in L\}$ and the similarity relation between possible worlds $\geq$. Here are some of the most interesting conditions (these conditions hold for all x, y, t, etc.; an A-world is a world in which A is true). In Rönnedal (2016) some of the relationships between these conditions are explored in detail.

(C – $c0$) If the set of A-worlds at t in M is the set of all B-worlds at t in M, then $R_A xyt$ iff $R_B xyt$.
(C – $c1$) If $R_A xyt$, then A is true in y at t in M.
(C – $c2$) If $R_A xyt$ and B is true in y at t in M, then $R_{A \wedge B} xyt$.
(C – $c3$) If the set of A-worlds at t in M is non-empty, then there is a y such that $R_A xyt$.
(C – $c4$) If $R_A xyt$ and B is true in y at t in M and $R_{A \wedge B} xzt$, then $R_A xzt$ and B is true in z at t in M.
(C – $c5$) If A is true in x at t in M, then $R_A xxt$.
(C – $c6$) If $R_A xyt$ and A is true in x at t in M, then $x = y$.
(C – $c7$) If $R_A xyt$ and $R_A xzt$, then $y = z$.

Table 1

(C – $cc2$) (($x \geq_w y$ or $y \geq_w x$) or ($x \geq_w y$ and $y \geq_w x$)).
(C – $cc3$) If the set of A-worlds at t in M is non-empty, then for all x there is a y such that A is true in y at t in M
    and if A is true in z at t in M then $y \geq_x z$.
(C – $cc4$) If $x \geq_w y$ and $y \geq_w z$, then $x \geq_w z$.
(C – $cc5$) $x \geq_x y$.
(C – $cc6$) If $y \geq_x x$, then $y = x$.
(C – $cc7$) If $y \geq_x z$ and $z \geq_x y$, then $y = z$.

Table 2

Proof theory

Tableau rules

Table 3 in Rönnedal ([2016](#)) includes the rules for $\boxdot$ and $\diamondsuit$. Since we do not use these rules in the present paper they are omitted here.

($\square \rightarrow$)	($\diamondsuit \rightarrow$)	($\neg \square \rightarrow$)	($\neg \diamondsuit \rightarrow$)
$A \square \rightarrow B, w_i t_k$	$A \diamondsuit \rightarrow B, w_i t_k$	$\neg(A \square \rightarrow B), w_i t_k$	$\neg(A \diamondsuit \rightarrow B), w_i t_k$
$r_A w_i w_j t_k$	$\downarrow$	$\downarrow$	$\downarrow$
$\downarrow$	$r_A w_i w_j t_k$	$A \diamondsuit \rightarrow \neg B, w_i t_k$	$A \square \rightarrow \neg B, w_i t_k$
$B, w_j t_k$	$B, w_j t_k$		
	where w_j is new		

Table 4

(CUT)	CId (CIdI)	CId (CIdII)
*	$\alpha(w_i)$	$\alpha(w_i)$
$\swarrow \searrow$	$w_i = w_j$	$w_j = w_i$
$\neg A, w_i t_k$ $A, w_i t_k$	$\downarrow$	$\downarrow$
for every A	$\alpha(w_j)$	$\alpha(w_j)$

Table 5

$T - c0$	$T - c0'$
If D is of the form $\boxdot (A \leftrightarrow B) \rightarrow$ $((A \,\square\!\rightarrow C) \leftrightarrow (B \,\square\!\rightarrow C))$, D, $w_i t_l$ can be added to any open branch on which $w_i t_l$ occurs	If D is of the form $((A \,\square\!\rightarrow B) \wedge (B \,\square\!\rightarrow A)) \rightarrow$ $((A \,\square\!\rightarrow C) \leftrightarrow (B \,\square\!\rightarrow C))$, D, $w_i t_l$ can be added to any open branch on which $w_i t_l$ occurs

Table 6

$T - c1$	$T - c2$	$T - c3$	$T - c4$
$r_A w_i w_j t_l$ ↓ A, $w_j t_l$	$r_A w_i w_j t_l$ B, $w_j t_l$ ↓ $r_{A \wedge B} w_i w_j t_l$	A, $w_i t_l$ ↓ $r_A w_j w_k t_l$ where w_k is new	$r_A w_i w_j t_l$ B, $w_j t_l$ $r_{A \wedge B} w_i w_k t_l$ ↓ $r_A w_i w_k t_l$ B, $w_k t_l$
$T - c5$	$T - c6$	$T - c7$	
A, $w_i t_l$ ↓ $r_A w_i w_i t_l$	A, $w_i t_l$ $r_A w_i w_j t_l$ ↓ $w_i = w_j$	$r_A w_i w_j t_l$ $r_A w_i w_k t_l$ ↓ $w_j = w_k$	

Table 7

(O)	(P)	$(\neg O)$	$(\neg P)$
OA, $w_i t_k$ $sw_i w_j t_k$ ↓ A, $w_j t_k$	PA, $w_i t_k$ ↓ $sw_i w_j t_k$ A, $w_j t_k$ where w_j is new	$\neg OA$, $w_i t_k$ ↓ $P\neg A$, $w_i t_k$	$\neg PA$, $w_i t_k$ ↓ $O\neg A$, $w_i t_k$

Table 8

The basic alethic and the basic temporal tableau rules are similar to the basic deontic rules that are included in Table 8. Finally, let us consider the rules for Rt, the propositional rules $(\neg \wedge)$ and $(\neg \leftrightarrow)$, and the rules $T - dD$, $T - MO$, $T - BT$ and $T - FT$ that are used in some tableaux in the paper (see Rönnedal 2012, for more on this).

(Rt)	$(\neg Rt)$	$(\neg \wedge)$	$(\neg \leftrightarrow)$
RtA, $w_i t_k$ ↓ A, $w_i t$	$\neg RtA$, $w_i t_k$ ↓ $Rt\neg A$, $w_i t_k$	$\neg(A \wedge B)$, $w_i t_k$ ↙ ↘ $\neg A$, $w_i t_k$ $\neg B$, $w_i t_k$	$\neg(A \leftrightarrow B)$, $w_i t_k$ ↙ ↘ A, $w_i t_k$ $\neg A$, $w_i t_k$ $\neg B$, $w_i t_k$ B, $w_i t_k$

Table 9

$(T - dD)$	$(T - MO)$	$(T - BT)$	$(T - FT)$
w_it_k	$sw_iw_jt_k$	A, w_it_k	A, w_it_k
$\downarrow$	$\downarrow$	$rw_iw_jt_k$	$rw_iw_jt_k$
$sw_iw_jt_k$	$rw_iw_jt_k$	$\downarrow$	$\downarrow$
where w_j is new		A, w_it_k	A, w_it_k
		where A is atomic	where A is atomic

Table 10

Tableau systems, the logic of a system, etc.

A semantic tableau system is a set of semantic tableau rules and the logic of a tableau system is the set of sentences (in the language L) that are theorems in this system. A sentence is a theorem in a system iff there is a proof of this sentence in the system. The concepts of proof, derivation, inconsistency, etc. are defined in a standard way.

In this paper, I use a special system called Strong Counterfactual Temporal Alethic-Deontic Logic (**SCTADL**) (see Section 3). This system includes all basic alethic, deontic and temporal rules, all rules in Table 3 – Table 6 and all rules in Table 7 except $T - c7$ in Rönnedal (2016; see above), all alethic rules that correspond to the fact that the alethic accessibility relation is 'reflexive', 'symmetric' and 'transitive', all deontic rules that correspond to the fact that the deontic accessibility relation is 'serial', 'transitive' and 'Euclidean', and all temporal rules that corresponds to the fact that the temporal accessibility relation is 'transitive', 'comparable' and does not branch towards the future or the past. This system is sound with respect to its corresponding class of models, that is, the class of models where the alethic accessibility relation is reflexive, symmetric and transitive, the deontic accessibility relation is serial, transitive and Euclidean, etc. (see Rönnedal 2016).

References

Åqvist, L. (1967). Good Samaritans, contrary-to-duty imperatives, and epistemic obligations. *Noûs, 1*, 361–379.

Åqvist, L. (1984). Deontic Logic. In D. Gabbay and F. Guenthner (eds.) *Handbook of Philosophical Logic*, Vol. II, D. Reidel, pp. 605–714.

Åqvist, L. (1987). *Introduction to deontic logic and the theory of normative systems*. Naples: Bibliopolis.

Åqvist, L. (2002). Deontic Logic. In Gabbay and Guenthner (Ed.), *Handbook of philosophical logic* (Vol. 8, 2nd ed., pp. 147–264). Dordrecht/Boston/London: Kluwer Academic Publishers.

Åqvist, L. (2003). Conditionality and Branching Time in Deontic Logic: Further Remarks on the Alchourrón and Bulygin (1983) Example. In Segerberg and Sliwinski (Ed.), *Logic, law, morality: thirteen essays in*

practical philosophy in honour of Lennart Åqvist, Uppsala Philosophical Studies (Vol. 51, pp. 13–37). Uppsala: Uppsala University.

Åqvist, L., & Hoepelman, J. (1981). Some theorems about a "tree" system of deontic tense logic. In R. Hilpinen (Ed.), *New studies in deontic logic* (pp. 187–221). Dordrecht: D. Reidel.

Bartha, P. (1993). Conditional obligation, deontic paradoxes, and the logic of agency. *Annals of Mathematics and Artificial Intelligence, 9*(1993), 1–23.

Belnap, N., Perloff, M., & Xu, M. (2001). *Facing the future: Agents and choices in our indeterminist world.* Oxford: Oxford University Press.

Bonevac, D. (1998). Against conditional obligation. *Noûs, 32*(Mars), 37–53.

Carmo, J. and Jones, A. J. I. (2002). Deontic logic and contrary-to-duties. In Gabbay and Guenthner (eds.) (2002) *Handbook of Philosophical Logic*, Vol. 8, pp. 265–343.

Castañeda, H. -N. (1977). Ought, time, and the Deontic Paradoxes. *Journal of Philosophy, 74*(12), 775–791.

Castañeda, H. -N. (1981). The paradoxes of deontic logic: The simplest solution to all of them in one fell swoop. In R. Hilpinen (Ed.), *New studies in deontic logic* (pp. 37–85). Dordrecht: D. Reidel.

Castañeda, H. -N. (1989). Paradoxes of moral reparation: Deontic foci vs circumstances. *Philosophical Studies, 57*, 1–21.

Chellas, B. F. (1969). *The logical form of imperatives*. Stanford: Perry Lane Press.

Chellas, B. F. (1980). *Modal logic: An introduction*. Cambridge: Cambridge University Press.

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

Cox, Azizah Al-Hibri. (1978). Deontic logic: A comprehensive appraisal and a new proposal. University Press of America.

Danielsson, S. (1968). *Preference and obligation: Studies in the logic of ethics*. Uppsala: Filosofiska föreningen.

DeCew, J. W. (1981). Conditional obligations and counterfactuals. *Journal of Philosophical Logic, 10*, 55–72.

Feldman, F. (1986). *Doing the best we can: An essay in informal deontic logic*. Dordrecht: D. Reidel Publishing Company.

Feldman, F. (1990). A simpler solution to the paradoxes of deontic logic. *Philosophical Perspectives, 4*, 309–341.

Forrester, J. W. (1984). Gentle murder, or the Adverbial Samaritan. *Journal of Philosophy, LXXI*(4), 193–197.

Gabbay, D., Horty, J., Parent, X., van der Meyden, E. & van der Torre, L. (eds.). (2013). Handbook of deontic logic and normative systems. College Publications.

Goble, L. (2003). Preference semantics for deontic logic part I- simple models. *Logique et Analyse, 46*(183–184), 383–418.

Goble, L. (2004). Preference semantics for deontic logic part II- multiplex models. *Logique et Analyse, 47*(335–363).

Greenspan P. S. (1975). Conditional Oughts and hypothetical imperatives. Journal of Philosophy, Vol. 72, No. 10 (May 22), pp. 259–276.

Hansson, B. (1969). An analysis of some deontic logic. *Noûs* 3, pp. 373–398. Reprinted in Hilpinen, Risto (ed). 1971. Deontic logic: Introductory and systematic readings. Dordrecht: D. Reidel Publishing Company, pp. 121–147.

Hilpinen, R. (Ed.). (1971). *Deontic logic: Introductory and systematic readings*. Dordrecht: D. Reidel Publishing Company.

Hilpinen, R. (Ed.). (1981). *New studies in deontic logic: Norms, actions, and the Foundation of Ethics*. Dordrecht: D. Reidel Publishing Company.

Horty, J. (2001). *Agency and deontic logic*. Oxford: Oxford University Press.

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

Lewis, D. (1974). Semantic analysis for dyadic deontic logic. In S. Stenlund (Ed.), *Logical theory and semantical analysis* (pp. 1–14). Dordrecht, Holland: D. Reidel Publishing Company.

Loewer, B., & Belzer, M. (1983). Dyadic deontic detachment. *Synthese, 54*, 295–318.

McNamara, P. (2004). Agential obligation as non-agential personal obligation plus agency. *Journal of Applied Logic, 2*, 117–152.

Montague, R. (1968). Pragmatics. In R. Klibansky (Ed.), *Contemporary philosophy: Vol. 1: Logic and the foundations of mathematics* (pp. 102–122). Firenze: La Nuova Italia Editrice.

Mott, P. L. (1973). On Chisholm's paradox. *Journal of Philosophical Logic, 2*, 197–211.

Meyer, J. -J. C. (1988). A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic, 29*(1).

Niles, I. (1997). Rescuing the counterfactual solution to Chisholm's paradox. *Philosophia, 25*, 351–371.

Parent, X. (2008). On the strong completeness of Åqvist's dyadic deontic logic G. In R. van der Meyden & L. van der Torre (Eds.), *Deontic Logic in Computer Science, Deon* (Vol. 2008, pp. 189–202). Berlin Heidelberg: Springer-Verlag.

Parent, X. (2010). A complete axiom set for Hansson's deontic logic DSDL2. *Logic Journal of the IGPL, 18*(3), 422–429.

Powers, L. (1967). Some Deontic Logicians. *Noûs, 1*, 361–400.

Prakken, H., & Sergot, M. (1996). Contrary-to-duty obligations. *Studia Logica, 57*, 91–115.

Rescher, N. (1958). An axiom system for deontic logic. *Philosophical Studies, 9*, 24–30.

Rönnedal, D. (2012). Temporal alethic-deontic logic and semantic tableaux. *Journal of Applied Logic, 10*(2012), 219–237.

Rönnedal, D. (2016). Counterfactuals in temporal alethic-deontic logic. *South American Journal of Logic, 2*(1), 57–81.

Smith, T. (1994). Legal expert systems: Discussion of theoretical assumptions. University of Utrecht.

Spohn, W. (1975). An analysis of Hansson's dyadic deontic logic. *Journal of Philosophical Logic, 4*, 237–252.

Tomberlin, J. E. (1981). Contrary-to-duty imperatives and conditional obligations. *Noûs, 15*, 357–375.

Tomberlin, J. E. (1989). Deontic paradox and conditional obligation. *Philosophy and Phenomenological Research., 50*(1), 107–114.

Tomberlin, J. E. (1989b). Obligation, conditionals, and the logic of conditional obligation. *Philosophical Studies, 55*, 81–92.

van Eck, J. (1982). A system of temporally relative modal and deontic predicate logic and its philosophical applications. *Logique et Analyse*, Vol. 25, no. 99, pp. 249–290, and no. 100, pp. 339–381. Original publication, as dissertation, Groningen, University of Groningen, 1981.

van Fraassen, C. (1972). The logic of conditional obligation. *Journal of Philosophical Logic, 1*, 417–438.

van Fraassen, C. (1973). Values and the Heart's command. *Journal of Philosophy* LXX, pp. 5–19.

von Kutschera, F. (1974). Normative Präferenzen und bedingte Gebote. In Lenk, H., & Berkemann J. (eds.), (1974), *Normenlogik: Grundprobleme der deontischen Logik*. UTB, 414, Verlag Dokumentation, Pullach (near München), pp. 137–165.

von Wright, G. H. (1964). A new system of deontic logic. *Danish Yearbook of Philosophy, 1*, 173–182.