DEONTIC LOGIC, WEAKENING AND DECISIONS CONCERNING DISJUNCTIVE OBLIGATIONS

Michael J. SHAFFER

ABSTRACT: This paper introduces two new paradoxes for standard deontic logic (SDL). They are importantly related to, but distinct from Ross' paradox. These two new paradoxes for SDL are *the simple weakening paradox* and *the complex weakening paradox*. Both of these paradoxes arise in virtue of the underlying logic of SDL and are consequences of the fact that SDL incorporates the principle known as weakening. These two paradoxes then show that SDL has counter-intuitive implications related to disjunctive obligations that arise in virtue of deontic weakening and in virtue of decisions concerning how to discharge such disjunctive obligations. The main result here is then that theorem T1 is a problematic component of SDL that needs to be addressed.

KEYWORDS: deontic logic, weakening, disjunction, decision theory, obligation

1. Introduction

SDL is the standard system of deontic logic that was developed importantly by von Wright and others in the 20th century.[1] The typical axiomatization of SDL is as follows:

A1. All well-formed tautologies.

A2. $\mathbf{O}(p \rightarrow q) \rightarrow (\mathbf{O}p \rightarrow \mathbf{O}q)$.

A3. $\mathbf{O}p \rightarrow \neg\mathbf{O}\neg p$.

R1. If p and $p \rightarrow q$, then q.

R2. If p, then $\mathbf{O}p$.[2]

So, SDL is just the modal logic KD with the operator **O** interpreted as "obligation" and **P** as "permission" such that **O**p is equivalent to $\neg\mathbf{P}\neg p$. From these axioms the following theorem importantly follows:

T1. $\mathbf{O}p \rightarrow \mathbf{O}(p \vee q)$.

[1] See von Wright 1951; McNamara 2019; and Hilpinen 2001.
[2] See Hilpinen 2001 and McNamara 2019.

Notice that T1 is a form of weakening that is embedded in SDL. Let us then begin to consider some important and problematic implications of SDL that arises in virtue of T1.

Suppose it is true that Op. By T1 it follows that O(p $\vee$ q). But note that this holds for *any and every* q_i whatever and it is reasonable to suppose that there are an infinite number of such alternatives or at least vastly more such alternatives than any agent could ever consider or even be aware of.[3] This may not seem to be immediately or obviously problematic, but it indeed is.[4] As we shall soon see, this is manifestly the case with respect to our intuitive and quite commonplace notion of disjunctive obligation and in the context of *decisions* with respect to the satisfaction of disjunctive obligations.

Let us suppose that a particular agent has the obligation to p and by T1 this implies that the agent also then has the obligation to do p or one of the large or infinite number of alternatives to p. For example, suppose Jane has an obligation to meet her friend Bonny for lunch on Tuesday. From this it follows that Jane has a disjunctive obligation to meet her friend Bonny for lunch on Tuesday or to meet her friend Bonny for lunch on Wednesday or to meet her friend Marge on Tuesday, etc., etc. In such a case it is reasonable to assume that such an agent can satisfy this sort of derived disjunctive obligation by doing p or any one of the alternatives to p and that the agent is *not* obligated to do each of the alternatives. But, this is at odds with SDL. Moreover, as we will see satisfying disjunctive obligations ultimately requires *choosing* to do p or one of the alternatives to p. However, given T1, it turns out that it is impossible to do this rationally in the context of orthodox decision theory and the rational obligation it imposes on us concerning choices. So from Op, T1 and the framework of orthodox decision theory it will be shown that we can derive the perplexingly absurd conclusion that it is impossible to have the obligation to p by a sort of reductio (at least provided we maintain orthodox decision theory). Let us examine these lines of argumentation in more detail in what follows. First, let us begin by introducing Ross' paradox, as it importantly presages the new (but related) criticisms of SDL introduced here.

---

[3] This is simply a consequence of the fact that natural languages and even reasonably realistic artificial languages are constituted by a potentially infinite set of sentences. This is simply a result of the compositionality of such languages.

[4] The problems alluded to here are related to but importantly *different* from Ross' paradox, as originally presented in Ross 1941 and Ross' free choice permission paradox from the same paper. Ross' paradox has nothing directly to do with iterated applications of T1 and the additional problems this raises.

2. Ross' Paradox

In 1941 Alf Ross introduced a seminal problem for SDL. This has come to be known as Ross' Paradox.[5] The paradox arises in virtue of SDL and it crucially involves T1. The problem arises from pairs of claims like these:[6]

(M1) It is obligatory that the letter is mailed.

(M2) It is obligatory that the letter is mailed or the letter is burned.

The formal SDL analogs of these claims are, respectively:

(M3) $\mathbf{O}p$

(M4) $\mathbf{O}(p \vee q)$

By T1, M4 follows from M3 and it would seem to be the case that the obligation in M4 can be satisfied *by burning the letter*. But this is clearly unacceptable. From one's having an obligation to mail the letter it should not follow that such an obligation can be fulfilled by burning the letter, especially given the reasonable supposition that burning the letter is (by assumption) forbidden and the fact that burning the letter would make it impossible to mail it. Something is clearly then wrong with SDL in virtue of its incorporation of T1. Ross' paradox crucially involves T1 and thus it is clearly the case that this problem results from the general idea of logical weakening (which is a familiar and seemingly unproblematic principle of propositional logic). As we shall see, however, the sort of *deontic weakening* that gives rise to Ross' paradox also gives rise to additional problems that are importantly related to, though distinct from, Ross' paradox.

3. The Simple Weakening Paradox

Suppose that an agent has an obligation to do p and that SDL is correct. By T1 this entails that the agent in question has a derived obligation to do p or q. However, in such a case it does not appear to be true that one must do all of the disjunctive actions and it does not appear to be true that one must do any one of the disjunctive options in particular. Crucially, this is supported by what Alan Donagan tells us about disjunctive obligations when he explains that: "…from the fact that I have a duty to save either a or b, it does not follow that I have a duty to save a and a duty to save b" (Donagan 2014). But it does seem to be the case that one must do something in such cases of disjunctive obligation. The intuitive idea is then that one can discharge such an obligation by doing either action. However, an agent facing such an

[5] See Ross 1941.

[6] See Ross 1941, 62 and McNamara 2019.

obligation is neither obligated to do both of the options nor is she specifically obligated to do the one and specifically obligated to do the other. This is all seemingly correct and, on this intuitive basis, the following principle concerning disjunctive obligation appears *prima facie* to be correct:

P1. $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}p \;\&\; \neg\mathbf{O}q \;\&\; \neg\mathbf{O}(p \;\&\; q)$.[7]

Again, this is just the intuitively grounded claim that the obligation to do p or q entails that it is not obligatory to do p, that it is not obligatory to do q, and that it is not obligatory to do both p and q. So, it would seem to be true that in the case of a disjunctive obligation an agent can discharge such an obligation by doing any one of the disjunctive actions and one need do no more than one of the two options in order to do so. But, this generates an immediate contradiction. We have derived $\neg\mathbf{O}p$ from $\mathbf{O}p$ as $\neg\mathbf{O}p$ is a conjunct of the consequent of P1 that follows from $\mathbf{O}p$ and T1. Let us call this the simple weakening paradox.

3.1 The Complex Weakening Paradox

So, something must go here and one obvious and natural suggestion is that P1 is the culprit. So, despite its intuitive plausibility, one might be tempted to deny P1 and replace it with the following principle:

P2. $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}(p \;\&\; q)$.

This alternative–and rather less intuitive–stance concerning disjunctive obligation defuses the simple weakening paradox with respect to SDL by simply omitting the implications of the formulas $\neg\mathbf{O}p$ and $\neg\mathbf{O}q$ that appear in P1.[8] P2 is just the claim

---

[7] One might initially think that this principle is overly strong, but the argument given in this section will work using any related principle such that $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}p$. In ordinary usage, it seems clearly to be the case that disjunctive obligations very often, or perhaps even always, do not entail specific obligations with respect to their disjuncts. For example, one's being obligated to either go to Paris or London does not seem to imply one's being obligated to go to Paris. Similarly, one's being obligated to donate $10 either to charity 1 or to charity 2 does not seem to entail that one is obligated to give $10 to charity 1 and that one is obligated to give $10 to charity 2. On the basis of such examples and a plethora of ones like them it is difficult really to see how one might reject this, but this possibility will be explored subsequently nevertheless. That point aside, that there are any plausible examples where $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}p$ holds means that SDL implies a host of contradictions (for there are surely many such cases) and this also suggests the problem raised here is not merely some artifact of the SDL formalism. Rather, it is a problem that arises in virtue of what appears to be the inadequacy of the SDL analysis/explication of obligation.

[8] Consider what the denial of P1 and the affirmation of P2 involves. If $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}(p \;\&\; q)$ but it is neither the case that $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}p$ nor is it the case that $\mathbf{O}(p \vee q) \rightarrow \neg\mathbf{O}q$, then it follows that $\mathbf{O}(p \vee q) \rightarrow \mathbf{O}p \;\&\; \mathbf{O}q$. This seems totally at odds with out intuitive notion of disjunctive

that the obligation to do p or q entails only that it is not obligatory to do both p and q. So, it still seems to be the case that when faced with a disjunctive obligation an agent could discharge such an obligation by doing one of the disjunctive actions, but the agent is not obligated to do both. Since, given this line of thinking, P1 is being denied and P2 affirmed it is not the case that Op implies $\neg Op$ and the simple weakening paradox is eliminated. Thus it would appear to be the case that satisfying an obligation to p could be accomplished as per SDL without immediately falling prey to the simple weakening paradox by rejecting P1 and adopting P2 instead of P1.

However, this response to the simple weakening paradox does not capture Donagan's insight at all, it is rather artificial and this solution still leaves us with a related and considerably more difficult problem related to T1. First, it should be clear that rejecting P1 and replacing P1 with P2 does not capture the insight that having an obligation to do p or q seems to imply the lack of an obligation to p and the lack of an obligation to q.[9] It is also an entirely arbitrary solution in that the replacement of P1 by P2 is solely motivated here by paradox avoidance at the expense of the intuitively well-grounded Donagan insight. More importantly, even if we ignore these charges, an additional problem arises when we recognize that if one's legitimate obligation to p obeys the deontic weakening principle T1 and P2 but not P1 applies to disjunctive obligations, then for any p the obligation to p entails an additional obligation to do one of an infinite or very large disjunction of possible options. In infinitary contexts (i.e. those where there are infinite possible alternatives to p) having an obligation to p, by weakening, implies having an obligation to do one of p or an infinite number of alternatives to p. In realistic finitary contexts (i.e. those where there are merely finite but very many possible options to p) having an obligation to p, by weakening, implies having an obligation to do one of p or one of the very large number of possible alternatives to p, many, or even most, of which will be unknown to the agent. In both cases P2 implies that it is not the case that one must do all of the disjunctive options. So, one must be able to satisfy such an obligation in principle by either doing one of an infinite number

---

obligation as evinced by Donagan's quote referenced earlier and as discussed in the previous footnote. One may be obligated to do p or q without being at all obligated to specifically do p and without being specifically obligated to do q. This may just indicate that SDL runs together more than one distinct notion of obligation. It has also been suggested that in modal contexts disjunction and conjunction are not used in the same manner as they are in non-modal contexts and this may be related to some of the problems raised here. See, for example, Geurts 2005 and Zimmermann 2000.

[9] Again, what this may suggest is that there are two sense of obligation at work in standard deontic logic.

of alternatives or by doing one of a very large number of alternatives many of which will be unknown to the agent. But, as we shall see, neither of these things can be done rationally.

All of this essentially means that agents have bona fide *options* with respect to satisfying disjunctive obligations–at least provided the options are individually satisfiable in the sense that the agent can do each of them–and, as per T1, every non-disjunctive obligation implies disjunctive obligations. What is curious then is that satisfying any disjunctive obligation is then a matter of choice.[10] This broadly comports with orthodox decision theory and, according to orthodox decision theory, an agent faces a choice when she has options that are within her power to perform upon choosing. As Levi points out,

> Having a choice presupposes having options. Having the option to perform some action entails having the ability to perform the action upon choosing it. Hence, having a choice presupposes having abilities to perform various actions upon choosing them (Levi 1986, 47).

Let us suppose then that the ability to perform each option from among the options facing our agent who is disjunctively obligated is within that agent's power. It follows then that the agent with any such disjunctive obligation must, as a result of this obligation, *choose* to do one of $(p \vee q_1 \vee q_2 \vee q_3 \vee \ldots \vee q_i)$ and importantly this holds even where the set of i alternatives to p is infinite or large and includes many options unknown to the agent. To discharge a disjunctive obligation *rationally* then, one must be able to rationally choose from amongst the options. In virtue of this requirement one must, of course, be able to identify the option that is the rational option. According to orthodox decision theory the rational decision is the one that maximizes expected utility with respect to the options an agent faces (i.e. one should rationally choose the option with the greatest utility value). In other words, the action that one *ought to do* is the one that maximizes expected utility. From all of this it follows that in such a situation it must be possible for the agent to choose to do one of p or the alternatives to p, one ought to decide by applying MEU and one's choice is rational just in case it actually maximizes expected utility. In the infinitary case this means then that if the agent is to rationally discharge the obligation to p or q–where there are an infinite number of q-alternatives to p–the agent must be able to efficaciously choose from amongst p and the infinite number of alternatives to p

[10] The relationship between deontic logic and decision theory, or course, involves controversial and difficult issues. However, the important and close relationship between the two types of formalisms has been explored in various ways and in terms of various conceptions of deontic logic in, for example, Hansson 1997; Torre and Tan 1997; Fusco 2015; Dietrich and List 2017; and Cariani forthcoming.

and the agent must choose the utility maximizing option from that set of possible options.

But this is impossible in the context of orthodox decision theory and the theory of preference on which it is based. Standard utility theory is based on the idea that if an agent's preferences obey the axioms of preference theory, then they can be represented as a utility function that exhibits certain supposedly desirable algebraic features. These axioms are introduced on the basis of their supposed intuitive (i.e. a priori) plausibility. Let "x ≼ y" mean "x is weakly preferred to y", "x ≺ y" mean "x is strictly preferred to y" (i.e. x is weakly preferable to y but x is not indifferent relative to y) and "x ∽ y" mean "x is indifferent relative to y" (i.e. x is weakly preferred to y and y is weakly preferred to x). Let O_i, O_j and O_k represent distinct outcomes and **p**, **q**, **r**,… represent distinct probability values. Finally, let $u(O_i)$ be a function representing a real numbered valuation of O_i. Given these basic representations we can then represent a gamble with a probability p of winning O_1 and a probability q of winning O_2 as [**p**O_1, (1 – **p**) O_2]. In terms of these representations, the axioms are used to characterize what is intuitively taken to be rational preference orderings are as follows.[11] First we have the ordering axiom:

> (U1) The preference relation ≽ is a total ordering that is reflexive and transitive.

Second, we have the better prizes axiom:

> (U2) For a fixed probability, prefer the gamble with a greater prize.

Third, we have the better chances axiom:

> (U3) For a fixed prize prefer the gamble with a greater probability.

Fourth, we have the reduction of compound gambles axiom:

> (U4) Compound gambles are to be evaluated in terms of the probability calculus.

Finally, we have the Archimedean or Continuity axiom:

> (U5) For any outcome that is ranked between two others there is a gamble between the more preferred and less preferred outcomes such that the agent is indifferent between it and the outcome ranked in between the more preferred and less preferred outcomes.

Formally, in terms of ≼ these axioms can be presented as follows:

> (U1.0) For any O_i and O_j either $O_i \preccurlyeq O_j$ or $O_j \preccurlyeq O_i$,

[11] This is the standard presentation of this representation theorem and it closely follows Bartha 2007. See Resnik 1987 and Gaus 2008 as well.

(U1.1) For any O_i, $O_i \preccurlyeq O_i$.

(U1.2) For any O_i, O_j and O_k, if $O_i \preccurlyeq O_j$ and $O_j \preccurlyeq O_k$, then $O_i \preccurlyeq O_k$.

(U2) $O_i \preccurlyeq O_j$, iff, for any $0 \leq \mathbf{p} \leq 1$ and any O_k, $[\mathbf{p}O_k, (1-\mathbf{p})O_i] \preccurlyeq [\mathbf{p}O_k, (1-\mathbf{p})O_j]$ and $[\mathbf{p}O_i, (1-\mathbf{p})O_k] \preccurlyeq [\mathbf{p}O_j, (1-\mathbf{p})O_k]$.

(U3) If $O_i \preccurlyeq O_j$, then for any $0 \leq \mathbf{p}, \mathbf{q} \leq 1$, $\mathbf{p} \geq \mathbf{q}$ iff $[\mathbf{p}O_i, (1-\mathbf{p})O_j] \preccurlyeq [\mathbf{q}O_i, (1-\mathbf{q})O_j]$

(U4) For any O_i and O_j and $\mathbf{p}, \mathbf{q}, \mathbf{r}$ such that $0 \leq \mathbf{p}, \mathbf{q}, \mathbf{r} \leq 1$, $[\mathbf{p}[\mathbf{q}O_i, (1-\mathbf{q})O_j], (1-\mathbf{p})[rO_i, (1-\mathbf{r})O_j]] \backsim [\mathbf{r}O_i, (1-\mathbf{t})O_j]$ for $\mathbf{t} = \mathbf{pq} + (1-\mathbf{p})\mathbf{r}$.

(U5) If $O_i \preccurlyeq O_j$ and $O_j \preccurlyeq O_k$, then there is a $\mathbf{p}$ such that $0 \leq \mathbf{p} \leq 1$ and $O_j \backsim [\mathbf{p}O_i, (1-\mathbf{p})O_k]$.

If an agent's preferences satisfy these axioms then those preferences can be represented by a real valued utility function $u(O_i)$ obeying the following two important conditions:

(C1) $O_i \preccurlyeq O_j$ iff $u(O_i) \leq u(O_j)$.

(C2) $u([\mathbf{p}O_i, (1-\mathbf{p})O_j]) = \mathbf{p}u(Oi) + (1-\mathbf{p})u(O_j)$.

The Expected Utility Theorem, the core idea behind utility theory, is then simply this claim that if one's preferences satisfy U1-U5, then those preferences can be represented as a real valued utility function satisfying C1 and C2.[12] In other words, formal utilities are a real-valued measure of preference and the value V_i of an outcome O_i is just $u(O_i)$ and according to the principle of maximizing expected utility (MEU)–the fundamental principle of rational decision making–one should rationally opt for the option with the highest utility value.

This second problem related to T1 appears to be that in SDL having any obligation to p implies having an obligation to do one of p or an infinite number of alternatives. This, in turn and in the context of orthodox decision theory, implies that one must be able to rationally make choices with respect to infinite sets of alternatives and according to orthodox decision theory we are obligated to maximize expected utility in such situations. But, this is impossible because one cannot rationally assign utilities to the outcomes of an infinite set of options. This is simply because one cannot have coherent decision-theoretic preferences with respect to an infinite set of outcomes that could be measured in terms of such utilities. This, in turn, is the case because one cannot have preferences with respect to an infinite set of outcomes. So, we have a deeply problematic paradox arising from the principles

[12] This is just the standard way of introducing utility theory via a representation theorem. This approach takes it as given a priori that U1-U5 are true. Recently, this approach to legitimizing decision theory has been challenged in Easwaran 2014 and by Meacham and Weisberg 2011.

of SDL, P2 and orthodox decision theory. Moreover, replacing P1 with P2 in order to avoid the simple weakening paradox that appeared to afflict SDL and P1 does nothing to help with this more complex paradox. Finally, even if we could circumscribe the set of alternatives to p in such cases and render the set of q-alternatives to p finite (say by arbitrarily assuming a finitary language involving a finite but very large set of alternatives for each p), the obligation to do p or one of a finite but very large number of q-alternatives to p–very many of which will be unknown to the agent–is impossible. Again, one cannot have coherent decision-theoretical utility assignments with respect to the outcomes of unknown alternatives. This is because an agent cannot have preferences with respect to such outcomes. So, this is not a viable response to the complex weakening paradox. One cannot maximize expected utility with respect to a choice involving unknown alternatives with unknown outcomes because one cannot have preferences with respect to unknown outcomes. So, we have our reductio of SDL. What should we make of all this? The most reasonable response seems to be that T1 is an objectionable principle. But, rejecting T1 requires revising SDL and it remains to be seen what such a deontic logic would look like.

4. Conclusion

SDL is a logical system based on modal logic that is supposed to capture the logic of permission and obligation, as well as other related concepts. So deontic logic is just the study of the logical properties of these notions. But SDL is limited in that it fails to capture at least some of our commonsense notions of these concepts. The catalog of deontic paradoxes that has been amassed since the inception of deontic logic shows this. Here two new paradoxes for SDL were introduced: the simple weakening paradox and the complex weakening paradox. Both paradoxes arise in virtue of the underlaying logic of SDL and are consequences of the fact that SDL incorporates the principle of weakening. These two paradoxes show that SDL has counter-intuitive implications related to disjunctive obligations that arise in virtue of deontic weakening and in virtue of decisions concerning how to discharge such disjunctive obligations. The main result here is then that theorem T1 is a problematic component of SDL that needs to be addressed.

References:

Bartha, P. 2007. "Taking Stock of Infinite Value: Pascal's Wager and Relative Utilities." *Synthese* 154: 5-52.

Cariani, F. Forthcoming. "Deontic Modals and Probabilities: One Theory to Rule Them All?" In *Deontic Modality*, edited by N. Charlow and M. Chrisman. Oxford University Press.

Dietrich, F. and C. List. 2017. "Choice-theoretic Deontic Logic." Manuscript: http://www.franzdietrich.net/Papers/DietrichList-ChoiceTheoreticDeonticLogic.pdf.

Donagan, A. 2014. "Consistency in Rationalist Moral Systems." *The Journal of Philosophy* 81: 291-309.

Easwaran, K. 2014. "Decision Theory without Representation Theorems." *Philosophers' Imprint* 14: 1-30.

Fusco, M. 2015. "Deontic Modality and the Semantics of Choice." *Philosophers' Imprint* 28: 1-27.

Gaus, G. 2008. *On Philosophy, Politics and Economics*. Belmont: Wadsworth.

Geurts, B. 2005. "Entertaining Alternatives: Disjunctions as Modals." *Natural Language Semantics* 13: 383–410.

Hansson, S. 1997. "Situationist Deontic Logic." *Journal of Philosophical Logic* 26: 423-448.

Hilpinen, R. 2001. "Deontic Logic." In *The Blackwell Guide to Philosophical Logic*, edited by L. Goble, 159-182. Malden: Blackwell Publishers.

Levi, I. 1986. *Hard Choices*. New York: Cambridge University Press.

Meacham, C. and J. Weisberg. 2011. "Representation Theorems and the Foundations of Decision Theory." *Australasian Journal of Philosophy* 89: 641-663.

McNamara, P. 2019. "Deontic Logic." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. URL = https://plato.stanford.edu/archives/fall2018/entries/logic-deontic/.

Resnik, M. 1987. *Choices: An Introduction to Decision Theory*. Minneapolis: University of Minneapolis Press.

Ross, A, 1941. "Imperatives and Logic." *Theoria* 7: 53–71.

Torre, L. and Y. Tan, "Prohairetic Deontic Logic and Qualitative Decision Theory." *AAAI 1997*: 1-9.

von Wright, G. H. 1951. "Deontic Logic." *Mind* 60: 1–15.

Zimmermann, T. 2000. "Free Choice Disjunction and Epistemic Possibility." *Natural Language Semantics* 8: 255–290.