

HELP FOR THE GOOD SAMARITAN PARADOX

(Received 18 October, 1985)

In a recent article James Forrester<sup>1</sup> formulates a novel version of the good samaritan paradox. Although there is considerable discussion of good samaritan paradoxes in the deontic logic literature, Forrester's paradox may appear to be especially intractable. Hector-Neri Castañeda has called it "the deepest paradox of deontic logic".<sup>2</sup> In this paper we will show how Forrester's paradox and other good Samaritan paradoxes can be resolved within a system of deontic logic 3-D which we have constructed and have applied to a number of issues concerning the logic of normative reasoning.<sup>3</sup>

1. GOOD SAMARITAN PARADOXES

Suppose that Arabella (the good Samaritan) ought to help Barbarella who has asked her to spare a dime. Now, Arabella will kill her husband's mistress next week and this person, unbeknownst to Arabella, happens to be Barbarella. Barbarella is the only person Arabella will kill next week. The paradox is that when these sentences are paraphrased into standard deontic logic (SDL) they seem to entail that Arabella ought to kill someone. The paradoxical argument in SDL is this: Let  $Hx$  and  $Kx$  stand for "Arabella helps  $x$ " and "Arabella kills  $x$  next week" and let  $b$  stand for "Barbarella". Then the paradoxical reasoning in SDL is as follows:

- (1)  $OHb$
- (2)  $b = !xKx$  (! is the definite description operator)

So

- (3)  $OH(!xKx)$
- (4)  $H(!xKx)$  entails  $ExKx$

So

- (5)  $O(ExKx)$

(5) follows from (3) and (4) in accordance with the principle

- (1) If  $p$  implies  $q$  then  $Op$  implies  $Oq$

which is valid in SDL. However, the argument is invalid in SDL since the step from (1) and (2) to (3) is invalid. Substitutivity of identity fails since “ $O$ ” is an intensional operator. The failure of substitutivity is easily explainable in terms of the usual possible world semantics for SDL. In these semantics “ $OA$ ” is true (at a world  $w$ ) iff “ $A$ ” is true at all the worlds that are permitted by the normative system (at  $w$ ). If  $OHb$  is true, then  $Hb$  is true at all the permitted worlds. (2) is insufficient for inferring (3) from (1) since it says only that  $b = !xKx$  holds in the actual world. A valid inference requires  $O(b = !xKx)$ , that  $b = !xKx$  holds in all the deontically permitted worlds.  $O(b = !xKx)$  is, of course, false since it is forbidden that Arabella kills anyone.

There is another way of formalizing the story so that substitutivity of identity holds:

- (1')  $Ex(x = b \ \& \ OHx)$

- (2')  $b = !xKx$

So

- (3')  $Ex(x = !xKx \ \& \ OHx)$

- (4)  $H(!xKx)$  entails  $ExKx$

So'

- (5)  $OExKx$

But this argument isn't valid either since (1) doesn't license an inference from (2'), (3') and (4) to (5). In terms of possible world semantics,  $b$  is the person in the actual world who Arabella will kill and in all deontically permissible worlds Arabella helps that person. But it doesn't follow that Arabella kills  $b$  (or anyone else) in all deontically permissible worlds. In fact,  $O - ExKx$  is consistent with (1') and (2').

The good samaritan paradox just discussed is easily disarmed by paying attention to scope distinctions. There are other versions of the paradox which cannot be dealt with in this way. The “paradox of the knower”<sup>4</sup> goes like this: If Arabella is going to commit a murder then Columbo, the detective, ought to know that she will. Suppose that Arabella is going to commit a

murder. It seems to follow that Columbo ought to know that Arabella will commit a murder. Up to this point the best representation of the story in SDL seems to be:  $m \rightarrow OKm$ ,  $m$ ; so  $OKm$  (“ $m$ ” is “Arabella will commit a murder” and “ $Km$ ” is “Columbo knows that Arabella will commit a murder”). But since  $Km$  entails  $m$ , it follows according to (I) that  $Om$ ; that Arabella ought to commit a murder. Can this paradox be resolved by appealing to scope distinctions? One suggestion is to understand “Columbo ought to know that Arabella will commit a murder as “ $m \& OBm$ ”; where “ $OBm$ ” means that Columbo ought to justifiably believe that Arabella will commit the murder. The strategy is to take  $m$  out of the scope of  $O$ . But  $OBm$  is almost as unacceptable as  $Om$ . In any case Columbo ought to know that Arabella will commit the murder if she will, not merely justifiably believe it. There does not seem to be a resolution of the paradox of the knower within SDL that depends on making scope distinctions.

Forrester’s paradox also seems to defy resolution in terms of scope distinctions. Forrester asks us to consider the following two ethical prescriptions:<sup>5</sup>

- (6) It ought to be that Smith not murder Jones.
- (7) It ought to be that if Smith murders Jones, Smith murders Jones gently.

He suggests that (6) and (7) should be paraphrased respectively by:

$$(6') \quad O - m$$

and

$$(7') \quad O(m \rightarrow g)$$

Forrester assumes the principle:

$$(C) \quad O(p \rightarrow q) \rightarrow (p \rightarrow Oq)$$

(C) is not valid in SDL. Adding it to SDL as an axiom results in the validity of  $p \rightarrow Op$ . This follows since an instance of (C) is  $O(p \rightarrow p) \rightarrow p \rightarrow Op$ . The antecedent of this conditional is valid in SDL so we obtain  $p \rightarrow Op$ . This is obviously undesirable. It is interesting to note that a modified version of (C) does hold in Castañeda’s system, though one that does not entail absurdities.<sup>8</sup> In any case, Forrester appeals to (C) to license the inference from (7’) to:

$$(8') \quad m \rightarrow Og$$

Now, suppose that Smith murders Jones,  $m$ . It follows that  $Og$ . But  $g$  entails  $m$  so by employing (I) we conclude  $Om$ . It looks as though a plausible story when formalized in SDL yields a very implausible consequence. The situation is even worse since  $O - m$  is part of the story and  $Om$  and  $O - m$  are inconsistent in SDL.

Since Forrester's paradox makes use of (C) which is not valid in SDL, one might think the "paradox" is not a paradox for proponents of SDL even though it may be one for Castañeda since the relevant instance of (C) does hold in his system. However, Forrester thinks the paradox is equally telling against SDL since even if (C) is not generally valid it does hold in this case. He imagines the following soliloquy on Smith's part:

"I am going to kill Jones; that much is settled. The only question is how. Now I don't want to infringe any more duties than is absolutely necessary. And as I read the law, if I do murder Jones then I ought to do so gently. And I will."<sup>6</sup>

Forrester's view is that once it is settled that Smith is going to kill Jones, he is under the obligation to do it gently. So even though (C) does not hold in general, in this case it is correct to infer  $Og$  from  $m$  and  $O(m \rightarrow g)$ . Forrester doesn't suggest any principles for determining when an unconditional obligation can be inferred from a conditional obligation and the truth of its antecedent. We will later show how 3-D provides an account of such inferences.

Walter Sinnott-Armstrong has recently suggested that Forrester's paradox like the good samaritan paradox discussed at the beginning of this paper rests on scope confusions.<sup>7</sup> He employs Donald Davidson's account of the logical form of action sentences in giving the following paraphrase of (7).

$$(7^*) \quad ExMxsj \rightarrow Ex(Mxsj \ \& \ OGx)$$

This says that if there is an action which is a murdering by Smith of Jones, then there is such an action and it ought to be gentle.  $ExMxsj$  and (7\*) do not imply  $OExMxsj$  (there ought to be a murder by Smith of Jones) though they do imply that there is an action which ought to be gentle. The suggestion is to remove the characterization of the action as a murder from the scope of the deontic operator. But this "solution" is not correct. It depends on analysing "Smith murders Jones gently" as  $Ex(Mxsj \ \& \ Gx)$  where "gently" is represented as a predicate of actions. But it is well known that certain adverbial expressions cannot be treated in this way. Suppose that Smith murders Jones by shoving him into the ice box where he freezes to death.

On Davidson's account Smith performed an action which was a shoving and also a murder. As a murder it was gentle (death by freezing is allegedly not as unpleasant as other deaths) but an especially ungentle shove might have been required to get Jones into the ice box. On Sinnott-Armstrong's account the murder would be both gentle and not gentle. Adverbial modifiers of action descriptions cannot always be treated as predicates of Davidsonian actions. This shows that Sinnott-Armstrong's account of the logical form of "Smith murders Jones gently" must be mistaken. If so, then there is no room for making the scope distinction on which his resolution of the paradox relies.

## 2. CASTAÑEDA'S TREATMENT OF THE GOOD SAMARITAN PARADOXES

Hector Castañeda has claimed that most of the paradoxes of deontic logic including good samaritan paradoxes can be resolved in his systems.<sup>8</sup> These systems differ from SDL in a number of ways. The most important involves a distinction he makes between propositions and practitions. Propositions are expressed by indicative sentences while practitions are expressed by infinitival expressions e.g. "Smith to kill Jones gently". In deontic contexts practitions describe the actions to be performed while propositions describe the circumstances of the action. For example, in "It ought to be that if Smith kills Jones, he kills him gently" the phrase "Smith kills Jones" expresses a proposition specifying circumstances in which the practition expressed by "he kills him gently" is to be satisfied. Deontic operators ( $O$  and  $P$ ) apply to practitions to form propositions. In Castañeda's system deontic contexts are extensional.  $O(a$  to help  $b)$  and  $b =$  the person  $a$  will kill next week entail  $O(a$  to help the person  $a$  will kill next week). Finally, the distinction between propositions and practitions results in the following modified versions of (I) and (C):

- (CI) If  $A$  implies  $B$  then  $OA$  implies  $OB$ .  
 (CC)  $O(p \rightarrow A) \rightarrow \cdot (p \rightarrow OA)$ .

Observe that in Castañeda's system (7) is formalized as  $O(m \rightarrow G)$  (" $m$ " is the proposition that Smith murders Jones and " $G$ " is the practition Smith to murder gently) and so does imply  $m \rightarrow OG$ . Since  $OG$  implies  $OM$  (" $M$ " is the practition Smith to murder Jones) Forrester's paradox arises in full

force in Castañeda's deontic systems. As Castañeda remarks, "... for a philosopher who holds that in general "ought (if  $p$ ,  $X$  to do  $A$ )" is equivalent to "if  $p$  then (ought ( $X$  to do  $A$ ))" the pressure of Forrester's paradox seems to be utterly suffocating. This is exactly my situation".

Castañeda's system is part of a comprehensive theory of deontic and practical reasoning.<sup>9</sup> We cannot discuss it in detail here but we do want to briefly explain how he handles good samaritan paradoxes including Forrester's paradox. Since deontic contexts are extensional in Castañeda's system he cannot use the resolution of the first good Samaritan paradox discussed in our paper. Instead he claims that attention to the distinction between propositions and practitions will provide solutions to both that paradox and the paradox of the knower. The basic idea is that the practition *Arabella to help the person she will kill next week* entails the proposition that Arabella will kill someone next week but it does not entail the practition *to kill someone next week*. Castañeda's proposition/practition distinction has the effect of separating a description of the circumstances of an action from a description of the action. The  $O$  operator applies only to the latter. Thus Castañeda's solution is similar to the second of the SDL solutions to the paradox since it in effect removes the statement that Arabella will kill someone from the scope of  $O$ . Although the paradox of the knower does not seem to be solvable by attention to scope distinctions, it does succumb to the proposition/practition distinction. "Columbo ought to know that Arabella will commit a murder" is represented by  $O$ (Columbo to know that Arabella will commit a murder). According to Castañeda the practition *Columbo to know that Arabella will commit a murder* does not entail the practition *Arabella to commit a murder* so (CI) cannot be employed to conclude that Arabella ought to commit a murder.

It should be clear that Forrester's paradox poses a difficult problem for Castañeda since it does not seem possible to separate practition from proposition in "Smith ought to murder Jones gently". Any action which is a gentle murder is a murder so the practition *Smith to murder Jones gently* entails the practition *Smith to murder Jones*. The step from  $OG$  to  $OM$  is not blocked merely by making the proposition/practition distinction.

Castañeda does offer a solution to Forrester's paradox. His idea is to distinguish between what he calls "core actions" and "aspectual actions". According to his view "Smith ought to murder Jones gently" represents two actions: the core action of murdering Jones and the aspectual action of

doing it gently. The deontic operator can apply to practitions corresponding to both actions or only to one. So, in effect, the sentence “Smith ought to murder Jones gently” is trebly ambiguous. In Forrester’s example “ought” applies only to the aspectual action (doing it gently) while the core action (murdering Jones) is part of the circumstances. The appropriate reading then is

- (J) “Smith will commit a murder of Jones and he ought to do it gently”.

Castañeda’s account seems to work as long as the practition “Smith to do it gently” does not entail the practition *Smith to murder Jones*. He thinks that there is no such entailment. But is this clear? How can one comply with the practition *to do it gently* as it occurs in (J) without murdering Jones? If one cannot, then it seems to follow that the practition *Smith to do it gently* does entail the practition *Smith to murder Jones*. Even if Castañeda is able to provide semantics for ought statements in which the entailment does not hold, his account depends on his view that when Jones murders Smith gently he does two things: murders Smith and does so gently. Although there may be arguments for this view (Castañeda gives some reasons for it other than its role in his resolution of Forrester’s paradox) it is bound to strike some philosophers as controversial. So it is interesting to see that the good samaritan paradoxes can be solved without this distinction in a system which is an extension of SDL.

### 3. THE 3-D SOLUTION TO THE PARADOX

We have constructed a system of deontic logic 3-D which we will show can resolve Forrester’s and the other good samaritan paradoxes. Our account of 3-D in this paper will be brief and informal.<sup>10</sup> We suppose that an ethical system (or any system of norms) induces a ranking on possible histories with respect to the extent to which the histories comply with the ethical system. The highest ranking possible histories are those at which no norm is violated. As one descends the ranking more and/or more serious violations occur. This allows for the evaluation of conditional obligation sentences.  $O(B/A)$  holds iff  $B$  holds at all the highest ranked histories at which  $A$  holds. If  $A$  holds then  $O(B/A)$  can be thought of (at first approximation) as expressing that  $A$  is a reason for making it the case that  $B$ . If  $A$  is true then  $B$  is a *prima facie*

obligation since there may be some other reason  $C$  and rule  $O(-B/A \ \& \ C)$  which defeats this obligation. In addition to prima facie obligations there are unconditional "all things considered" obligations. On our account what a person ought to do all things considered changes over time as the facts to be considered changes. I ought today to finish a paper by September. But if next week I receive a letter from the editor extending the deadline to November that obligation disappears (and is replaced by a new one).

At time  $t$  at history  $h$  there are typically many futures which are possible. We will say that these histories extend  $h$  at  $t$ . What ought to be all things considered according to the value system at  $h$  at time  $t$  is characterized by the set of highest ranked histories that extend  $h$  at  $t$ . If  $B$  is true at all such histories than  $O_t B$  is true at  $h$ .

We can develop the preceding ideas to make a connection between conditional ought statements and all things considered ought statements. Suppose that  $O(B/A)$  and  $A$  are true at  $h$ . Can we conclude that  $O_t B$  holds for some (or all) times at  $h$ ? Clearly not. First of all, at time  $t$  there may be extensions of  $h$  at which  $-A$  and  $-B$  are true which are ranked more highly than any history at which  $A$  holds. So  $O_t -A$ . To avoid this we need to require that  $A$  is true at all the histories which extend  $h$  at  $t$ . When this holds we will say that  $A$  is settled at time  $t$  (at history  $h$ ),  $S_t A$ . We also need to assume that  $B$  is possible at  $t$  (i.e.  $-B$  is not settled at  $t$ ). But this isn't sufficient since there may be some other value statement e.g.  $O(-B/A \ \& \ C)$  which holds. If  $A \ \& \ C$  is settled at  $t$  then given  $O(-B/A \ \& \ C)$ ,  $O(B/A)$  is defeated. However, we can infer  $O_t B$  from  $O(B/A)$ ,  $S_t A$ ,  $-S_t -B$ , and a statement which says that there is no true value statement  $O(-B/A \ \& \ C)$  which defeats  $O(B/A)$  at  $t$ . We will abbreviate the last clause by  $U_t(B, A)$ . Truth conditions for  $U_t(B, A)$  can easily be given in terms of the model of branching histories ranked by a value system. The following inference is valid in 3-D:

$$\begin{array}{l}
 (3d) \quad O(B/A) \\
 \quad -S_t -B \\
 \quad S_t A \\
 \quad \underline{U_t(B, A)} \\
 \quad O_t B
 \end{array}$$

This inference allows one to infer an unconditional obligation from conditional obligation statements.

We will now represent Forrester's paradox in 3-D. We will suppose a simple



value system whose only rules (these are all that are relevant to Forrester's story) are  $O(-m/T)$  and  $O(g/m)$  ( $T$  is a tautology). The first says that in the most highly ranked worlds (since  $T$  is true in all worlds) Smith does not murder Jones. The second says that in the highest ranking of those (not very good) worlds at which Smith does murder Jones he does so gently. At time  $t$  before it is settled that Smith murders Jones the value system prescribes that Smith ought not murder Jones;  $O_t - m$  is true. At  $t'$  when it is settled that Smith will murder Jones, it ought to be that Smith murders Jones gently since  $g$  is true at the most highly ranked histories which extend the actual history at  $t'$ .

The preceding analysis goes some distance towards resolving Forrester's paradox. We have avoided the unfortunate consequence of the SDL representation of Forrester's story that both  $Om$  and  $O - m$  are true (a contradiction in SDL) by recognizing that what unconditionally ought to be changes over time. Furthermore, our reconstruction of the reasoning which leads to  $O_t g$  is more satisfactory than Forrester's which employs the invalid (C). In our representation the role of (C) is played by the inference scheme (3d) which is valid in 3-D.

An objection to our analysis is that although Smith has decided to murder Jones by  $t'$  it is not settled that he will in our sense of settled. He could still change his mind. But if this is so then we would maintain that Smith does not have at  $t'$  the obligation to murder Jones gently. He has the obligation not to murder Jones at all although, of course, the conditional statement  $O(g/m)$  remains true.

We must admit that we have yet not fully resolved Forrester's paradox. The problem is that  $O_t g$  implies  $O_t m$ . Even if there is no contradiction it may still seem incorrect that at  $t'$  it ought to be that Smith murders Jones. The solution is suggested by the observation that the reason that  $O_t m$  holds is not that the value system recommends murder but that at  $t'$   $m$  is settled. One of the features of 3-D is that  $S_t p$  implies  $O_t p$ . We might say that 3-D makes a virtue (or rather an obligation) of necessity. When one remembers that the point of the practical reasoning expressible in 3-D is to guide deliberation this may not seem so surprising. After all, there is no point in deliberating about what is settled. In any case we can avoid whatever counterintuitiveness one may find in  $O_t m$  by defining another deontic operator  $O_t^* p$  as holding whenever  $O_t p$  and  $\neg S_t p$  hold.<sup>11</sup> Principle (I) obviously does not hold for  $O_t^* p$ . But a slight modification does hold:

(I\*) If  $p$  entails  $q$  then  $\neg S_t q$  and  $O_t^* p$  entails  $O_t^* q$ .

Forrester's paradox can now be completely solved. Although we can derive  $O_t^* g$  from the premises of Forrester's story we cannot derive  $O_t^* m$ . The fact that  $m$  is settled at  $t$  is needed to derive  $O_t^* g$  but that very fact makes  $O_t^* m$  false.

The paradox of the knower is similarly resolved in 3-D. "If Arabella will commit a murder then Columbo ought to know it" is represented by  $O(Km/m)$ . If  $O(Km/m)$  is not defeated at  $t$  and if it is settled at  $t$  that  $m$  then we can conclude  $O_t^* Km$ . But we cannot conclude  $O_t^* m$  since  $m$  is settled at  $t$ .

While others have suggested that the way to solve the good samaritan paradoxes is to drop (I)<sup>12</sup> our solution shows exactly how (I) should be modified. In addition, we are able in 3-D to analyse the relationship between conditional obligations and all things considered obligations. Instead of the very questionable principle (C) we have the inference scheme (3d) which we maintain captures what is correct about the view that one can sometimes infer all things considered ought statements from conditional ought statements. We can also shed some light on Castañeda's claim that deontic contexts are extensional. Generally, the inference from  $O_t Fa$  and  $a = b$  to  $O_t Fb$  is invalid in 3-D. However, if  $a = b$  is settled at  $t$  then the resulting inference is valid. So Castañeda's claim is partially captured in 3-D.

As Castañeda maintains, philosophical systems and, more relevantly, deontic logics should be compared with respect to the whole range of problems they are supposed to handle. It would be interesting and instructive to provide such a comparison between 3-D and Castañeda's systems. While we have not provided such a broad comparison we do think that we have shown that 3-D can deal with good Samaritan paradoxes at least as well as Castañeda's system.

#### NOTES

<sup>1</sup> Forrester, James, 'Gentle murder, or the adverbial Samaritan', *The Journal of Philosophy* LXXXI, 4 (April 1984).

<sup>2</sup> Castañeda, Hector-Neri, 'Aspectual actions and the deepest paradox of deontic logic', forthcoming in papers from the Davidson Conference, ed. E. Lepore. New Brunswick N.J. (April 30, 1984).

<sup>3</sup> Loewer, Barry and Belzer, Marvin, 'Dyadic deontic detachment', *Synthese* 54 (1983).

<sup>4</sup> Aqvist, Lennart, 'Good Samaritans, contrary to duty imperatives and epistemic obligations', *Noûs* I, 4 (December 1967).

<sup>5</sup> Forrester op. cit.

<sup>6</sup> Ibid. p. 196.

<sup>7</sup> Sinnott-Armstrong, Walter, 'A solution to Forrester's paradox of gentle murder', *The Journal of Philosophy* LXXXII, 3 (March 1985).

<sup>8</sup> Castañeda, Hector-Neri, 'The paradoxes of deontic logic', in Hilpinen, R. (ed.) *New Studies in Deontic Logic*, D. Reidel, 1981.

<sup>9</sup> Castañeda, Hector-Neri, *Thinking and Doing*, D. Reidel, 1975.

<sup>10</sup> Loewer and Belzer, cited in Note 3.

<sup>11</sup> Fred Feldman introduces an operator like out  $O^*$  in 'Obligations-absolute, conditioned, and conditional', *Philosophia* 12 (March 1983).

<sup>12</sup> Forrester op. cit. p. 196.

*Department of Philosophy,  
University of South Carolina,  
Columbia, SC 29208,  
U.S.A.*

*Department of Philosophy,  
University of Missouri,  
Columbia, MO 65211,  
U.S.A.*