

# A Note on the Good Samaritan Paradox and the Disquotation Theory of Propositional Content

Steven Orla Kimbrough  
University of Pennsylvania  
Philadelphia, PA 19104 USA  
kimbrough@wharton.upenn.edu

## 1 The Paradox

Standard deontic logic (SDL) contains the following rule.

**Expression 1** *If  $\vdash \phi \rightarrow \psi$ , then  $\vdash \mathcal{O}\phi \rightarrow \mathcal{O}\psi$*

This rule is also known as the good Samaritan Paradox, since it fits (with allowance for stylistic variance) the form:

$\phi$ : The good Samaritan helps the victim, who has been hurt.

$\psi$ : The victim has been hurt.

But then it follows—given that it is obligatory to help a victim who has been hurt—that it is obligatory that  $\psi$ , and surely that is paradoxical, if not absolutely wrong.

The Paradox of the Knower, or the Paradox of Epistemic Obligation [4] is a similar, if not identical, problem. Suppose:

1.  $\phi$
2.  $\mathcal{F}\phi$   
(It is forbidden that  $\phi$ .)
3. If  $\phi$ , then  $\mathcal{O}\mathcal{K}_i\phi$   
(If  $\phi$ , then [pick your favorite version] it is obligatory that  $i$  know that  $\phi$ .)

Now, (1) and (3) imply

4.  $\mathcal{O}\mathcal{K}_i\phi$

Add in the epistemic principle

5.  $\mathcal{K}_i\phi \rightarrow \phi$   
(If  $i$  knows that  $\phi$ , then  $\phi$  is true.)

and with Expression 1 you get

6.  $\mathcal{O}\phi$

Let  $\phi$  be “ $i$ ’s wife is committing adultery,” and we have a forceful example (to many non-swinging husbands at least) of a paradox.

## 2 Individuating Violation States

SDL isn’t the only form of mainline deontic logic that is subject to this paradox. It arises in the Anderson reduction [1, 7], and indeed in all other (mainline) forms I know of. But the Anderson reduction is a step in the right direction. Briefly, instead of  $\mathcal{O}\phi$  for “It ought to be the case that  $\phi$ ” we have  $\Box(\neg\phi \rightarrow V)$  where  $V$  is the bad (violation) condition. That is, “ $\phi$  ought to be true” is unpacked as “Necessarily, if  $\phi$  isn’t true, then the bad happens” (and that’s not good!).

Under the Anderson reduction every violation is equivalent in the sense that  $V$  obtains. Let us suppose (temporarily without justification) that instead every deontic condition (obligation, forbiddance, etc.) can be individuated in the sense that each is associated with a possibly unique violation,  $V(x)$ . Specifically, assume that for every statement  $\phi$

1.  $\mathcal{O}\phi \rightarrow \exists x(\neg\phi \leftrightarrow V(x))$
2.  $\mathcal{F}\phi \rightarrow \exists x(\phi \leftrightarrow V(x))$

(I assume that  $x$  does not occur freely in  $\phi$ .) Let us now revisit the paradox argument (above), substituting consequents for their antecedents.

1.  $\phi$
2.  $\exists x(V(x) \leftrightarrow \phi)$   
(From  $\mathcal{F}\phi$ , it is forbidden that  $\phi$ .)
3. If  $\phi$ , then  $\exists x(V(x) \leftrightarrow \neg\mathcal{K}_i\phi)$   
(From if  $\phi$ , then  $\mathcal{O}\mathcal{K}_i\phi$ ; if  $\phi$ , then [pick your favorite version] it is obligatory that  $i$  know that  $\phi$ .)

Now, (1) and (3) imply

4.  $\exists x(V(x) \leftrightarrow \neg\mathcal{K}_i\phi)$   
(Formerly,  $\mathcal{O}\mathcal{K}_i\phi$ )

Add in the epistemic principle

5.  $\mathcal{K}_i\phi \rightarrow \phi$   
(If  $i$  knows that  $\phi$ , then  $\phi$  is true.)

and *without* Expression 1 you no longer get

6.  $\mathcal{O}\phi$

Instead, you do get

7.  $\exists x(\neg\phi \rightarrow V(x))$

which looks very much like the Anderson reduction without the necessity operator affixed. But, the biconditional version of  $\mathcal{O}\phi \rightarrow \exists x(\neg\phi \leftrightarrow V(x))$  is not something we would wish to subscribe to. So, the implication to  $\mathcal{O}\phi$  is blocked.

For many reasons (including intensionality) this account will not do as it stands. Still, it can yield more insight before we abandon it in favor of another.

It is instructive to impose existential instantiation where permitted in the argument above. Here in brief is what happens.

1.  $\phi$
2.  $(V(a) \leftrightarrow \phi)$
3. If  $\phi$ , then  $(\neg\mathcal{K}_i\phi \leftrightarrow V(b))$

Now, (1) and (3) imply

4.  $(\neg\mathcal{K}_i\phi \leftrightarrow V(b))$

Add in the epistemic principle

5.  $\mathcal{K}_i\phi \rightarrow \phi$

and *without* Expression 1 you no longer get

6.  $\mathcal{O}\phi$

Instead, you do get

$$7. (\neg\phi \rightarrow V(b))$$

Points arising:

1. We are distinguishing two obligations: the obligation for spouse  $a$  not to commit adultery and the (conditioned) obligation that spouse  $b$  (or  $i$ ) know about the adultery of spouse  $a$ . This allows us to make sense of the conclusion, (7). It is simply another case of the paradox of material implication, something most (or many) are happy to live with. If the wife is committing adultery, then if the wife is not committing adultery, then [anything you want]. And note that  $(\neg\phi \rightarrow V(a))$  does not follow.<sup>1</sup>
2. We've done without Expression 1, but something similar obtains.

**Expression 2** *If  $\vdash \phi \rightarrow \psi$ , then  $\vdash \exists x(\neg\phi \leftrightarrow V(x)) \rightarrow \exists x(\neg\psi \rightarrow V(x))$*

I want to suggest that the intuition served by Expression 1 is also served—at least nearly as well and without the paradoxical consequences—by Expression 2. But I defer to another time an argument for this suggestion.

3. The approach just outlined is a prototype presented for the sake of exposition. It raises a great number of semantic and even metaphysical questions. I will have a little to say about these issues in the concluding section. First, let us see how the prototype might be embedded in an approach that promises to be a serious candidate.

### 3 Disquotation Theory

The disquotation theory [5] is an extension of event semantics [2, 3, 6, 8]. It offers a first-order representation for handling intensional contexts. In event, or subatomic, semantics, events, processes, states, and perhaps other kinds of entities are seen as fundamental constituents of verbs, as what the verbs in a sense are about. “Jane arrives,” for example, is analyzed (roughly) as a stylistic variant of “There is an arrival event,  $e$ , and Jane is the subject of  $e$ .” First-order representation often becomes quite straightforward. In the present case, roughly:

---

<sup>1</sup>That is,  $(\neg\phi \rightarrow V(b))$  follows from  $\mathcal{K}_i\phi \rightarrow \phi$  and  $\neg\mathcal{K}_i\phi \leftrightarrow (V(b))$ , but  $(\neg\phi \rightarrow V(a))$  does not. Both follow from  $\phi$ .

**Expression 3**  $\exists e(\text{arrive}(e) \wedge \text{Subject}(e, \text{Jane}))$ .

Similarly, “*a* delivers *g* to *s*” might be represented as

**Expression 4**  $\exists e_1(\text{deliver}(e_1) \wedge \text{Sub}(e_1, a) \wedge \text{Obj}(e_1, g) \wedge \text{IndObj}(e_1, s))$

since *a* is the subject, *Sub*, *g* is the direct object, *Obj*, and *s* is the indirect object, *IndObj*.<sup>2</sup>

The disquotation theory treats intensional verbs as taking direct objects that are specially quoted statements. For example, “Bob said that Jane arrived” might be represented as

**Expression 5**  $\exists e_1, t_1(\text{say}(e_1) \wedge \text{Subject}(e_1, \text{Bob}) \wedge \text{Obj}(e_1, [\exists e, t(\text{arrive}(e) \wedge \text{Subject}(e, \text{Jane}) \wedge \text{Cul}(e, t) \wedge t < \text{now})]) \wedge \text{Cul}(e_1, t_1) \wedge t_1 < \text{now})$

Further, one or more axiom schemas are associated with each verb. In the present example we would likely have an axiom schema to the effect that a saying (event) *e* is truthful, or veridical, *Ver(e)*, if and only if the object of the event, what was said, is itself true. This representation is achieved by removing the special quotation operators, or disquoting the propositional content of the sentence. For example:

**Axiom Schema 1 (Strong Say Rule)**

$\forall e((\text{say}(e) \wedge \text{Obj}(e, [\phi])) \rightarrow (\phi \leftrightarrow \text{Ver}(e)))$

The gist of the schema is that if you say that  $\phi$  then your saying is truthful, or veridical, if and only if  $\phi$ . Points arising:

1. The effect of the quotation operator,  $[\cdot]$ , is to map a well-formed formula into a logical name. This results in a very high degree of opacity. To deduce  $F([\phi])$  from  $F([p])$ ,  $p \leftrightarrow \phi$  is not sufficient, nor is  $\Box(p \leftrightarrow \phi)$ , nor is “*p* means  $\phi$ .” Only  $[p] = [\phi]$  will do.
2. The disquotation approach (in conjunction with use of event semantics) permits us to distinguish and identify every verb instance (that is modelled). The verb’s argument may be thought of as a globally unique identifier. See point 1 of the points arising at the end of the previous section.

---

<sup>2</sup>Having my druthers I would use semantic rôles instead of standard grammatical categories such as subject and indirect object, but this is a complication that adds little and invites confusion in the present context.

These all-too-brief remarks will have to suffice for limning the disquotatation theory. We now focus on its application to representation for deontic reasoning.

I have previously pointed out that the Anderson reduction may be re-framed so as to exploit the disquotatation theory (with underlying events) for representation of propositional content [5]. Suppose that a delivery is obligated:

**Expression 6**  $\mathcal{O}\exists e_1(\text{deliver}(e_1) \wedge \text{Sub}(e_1, a) \wedge \text{Obj}(e_1, g) \wedge \text{IndObj}(e_1, s))$

Our fundamental schema for *ought* has the form

**Fundamental Schema 1 (ought)**  $\exists e(\text{ought}(e) \wedge \text{Obj}(e, [\phi]) \wedge \Gamma)$

and our example (Expression 6) instantiates in the following way:

**Expression 7**  $\exists e(\text{ought}(e) \wedge \text{Obj}(e, [\exists e_1(\text{deliver}(e_1) \wedge \text{Sub}(e_1, a) \wedge \text{Obj}(e_1, g) \wedge \text{IndObj}(e_1, s)]))$

(Here and elsewhere,  $\exists e(\text{ought}(e)) \dots$  reads “There is a state,  $e$ , such that it is an obligation state. . .”. Corresponding closely to the spirit of the Anderson reduction gives us the *weak* ought rule:

**Axiom Schema 2 (Weak Ought Rule)**

$\forall e((\text{ought}(e) \wedge \text{Obj}(e, [\phi])) \rightarrow (\neg\phi \rightarrow V(e)))$

Note that we have  $(\neg\phi \rightarrow V(e))$  rather than  $\Box(\neg\phi \rightarrow V(e))$  as in the Anderson reduction. This is as it should be. The fundamental schema ensures sufficient intensionality, so that if  $\phi$  ought to be the case and  $\phi \leftrightarrow \psi$ , it does *not* follow merely from the fundamental schema that  $\psi$  ought to be the case. Moreover, it does follow from the axiom schema that if  $\neg\psi$  then the same violation condition obtains when  $\neg\phi$ .

Our use of event semantics permits distinguishing  $V$  more finely as  $V(e)$ ; instead of the *the* violation condition,  $e$  names *a* violation condition. This allows us to employ the *strong* ought rule:

**Axiom Schema 3 (Strong Ought Rule)**

$\forall e((\text{ought}(e) \wedge \text{Obj}(e, [\phi])) \rightarrow (\neg\phi \leftrightarrow V(e)))$

Permission works similarly.

**Expression 8 (Permission)**  $\exists e(\text{permit}(e) \wedge \text{Obj}(e, [\exists e_1(\text{deliver}(e_1) \wedge \text{Sub}(e_1, a) \wedge \text{Obj}(e_1, g) \wedge \text{IndObj}(e_1, s)]))$

**Fundamental Schema 2 (Permission)**  $\exists e(\text{permit}(e) \wedge \text{Obj}(e, [\phi]) \wedge \Gamma)$

Permissions don't lead to violations. You can't violate a permission.

**Axiom Schema 4 (Permission Rule)**

$\forall e((\text{permit}(e) \wedge \text{Obj}(e, [\phi])) \rightarrow \neg V(e))$

Finally, prohibition, for which we have weak and strong rules, as with obligation.

**Axiom Schema 5 (Weak Prohibition Rule)**

$\forall e((\text{prohibit}(e) \wedge \text{Obj}(e, [\phi])) \rightarrow (\phi \rightarrow V(e)))$

**Axiom Schema 6 (Strong Prohibition Rule)**

$\forall e((\text{prohibit}(e) \wedge \text{Obj}(e, [\phi])) \rightarrow (\phi \leftrightarrow V(e)))$

I draw the reader's attention to the strong ought rule, Axiom Schema 3. Let us recapitulate the Paradox of the Knower in light of the disquotation theory and the strong ought rule. Suppose:

1.  $\phi$
2.  $\mathcal{F}\phi$   
(It is forbidden that  $\phi$ . Now: Axiom Schema 6)
3. If  $\phi$ , then  $\mathcal{OK}_i\phi$   
(If  $\phi$ , then [pick your favorite version] it is obligatory that  $i$  know that  $\phi$ .)

Now, (1) and (3) imply

4.  $\mathcal{OK}_i\phi$

Add in the epistemic principle

5.  $\mathcal{K}_i\phi \rightarrow \phi$   
(If  $i$  knows that  $\phi$ , then  $\phi$  is true.)

and *without* Expression 1 you no longer get

6.  $\mathcal{O}\phi$

instead you have Axiom Schema 3 so all you get is

**Expression 9**

$$\forall e((ought(e) \wedge Obj(e, [\mathcal{K}_i\phi])) \rightarrow (\neg\mathcal{K}_i\phi \leftrightarrow V(e)))$$

and then

**Expression 10**

$$\forall e((ought(e) \wedge Obj(e, [\mathcal{K}_i\phi])) \rightarrow (\neg\phi \rightarrow V(e)))$$

We are assuming, of course,

**Expression 11**

$$\exists e(ought(e) \wedge Obj(e, [\mathcal{K}_i\phi]))$$

Note the similarity of Expression 10 to the Anderson reduction. With a little charity and backwards engineering the two might be taken as morally equivalent.

Let us look closer. Again, let  $\phi$  be “*i*’s wife is committing adultery.” What Expression 10 amounts to is that if *i*’s wife is *not* committing adultery then any  $e$  instantiating Expression 11 will be a violation state. Indeed, this is as it should be. Since  $\neg\psi \rightarrow \neg\mathcal{K}_i\psi$ , the governing obligation is violated. But that is all. It does *not* follow either that

**Expression 12**

$$\exists e((ought(e) \wedge Obj(e, [\phi])))$$

or that

**Expression 13**

$$\forall e((ought(e) \wedge Obj(e, [\mathcal{K}_i\phi])) \rightarrow (\neg\phi \leftrightarrow V(e)))$$

To get Expression 13 we would need to strengthen  $\mathcal{K}_i\psi \rightarrow \psi$  to  $\mathcal{K}_i\psi \leftrightarrow \psi$ , and we surely don’t want to do that.

The upshot, I suggest, is this. Deontic logic needs the strong ought rule, Axiom Schema 3, or something more or less equivalent. SDL and the Anderson reduction do not have it. Rather they have Axiom Schema 2, or something more or less equivalent.

Much remains to be investigated, but the disquotaton theory appears to have some attractive properties for formalizing deontic reasoning.

## 4 Discussion

In this discussion note, I have aimed to illustrate how the disquotaton theory (with event semantics) can provide a principled and attractive solution

to the Paradox of the Knower. §2 presented a stripped-down version of the disquotation theory, focusing on use of event semantics. There I argued that event semantics can permit us to distinguish among obligations (and violations of the obligations) and that this is an important key to unlocking the paradox. §3 presented a sketch of the disquotation theory. Its representational pattern generates the stripped-down formulas used in the exposition in §2. The promising findings and lessons of §2 are supported by the more thoroughgoing (but hardly complete) account of §3.

Very much remains to be done, explored, and explained. I close with two remarks in that direction.

1. I conjecture that Expression 1 is an unsatisfactory principle. Instead of obligation,  $\mathcal{O}$ , consider for a moment an operator for asserting,  $\mathcal{A}$ . Is the analog of Expression 1 acceptable? I think not. If I assert that  $\phi$  and it happens that  $\phi \rightarrow \psi$  is true (and provable), have I also asserted (should it be provable that)  $\psi$ ? Surely not. What should follow is that if  $\neg\psi$  then my assertion is not truthful (whatever my intentions happen to be). As argued elsewhere, the disquotation theory captures this intuition.

Can a similar intuition be sustained for obligation and other deontic concepts? I think so. If  $\phi$  is (provably) obligatory and (provably)  $\phi \rightarrow \psi$ , is  $\psi$  obligatory? Surely, if  $\neg\psi$  then there will be a violation of the obligation on  $\phi$ . Why go further and insist that  $\psi$  is also obligatory? The Paradox of the Knower can be interpreted as cautioning us not to take this next step.

2. What are we to make of  $x$  in  $\exists x(\text{oblige}(x) \dots)$ ? Can we instantiate such expressions and if so, what sort of thing is  $a$  in  $(\text{oblige}(a) \dots)$ ? These are large questions and justice demands a much fuller treatment than can be undertaken here. Briefly, however, there's this.  $a$  in  $(\text{oblige}(a) \dots)$  is a state of being an obligation. There are very many states, including being a socialist, being north of the equator, and so on. Being an obligation is just one kind of state. How do obligation states come into existence? It is far from clear that the disquotation theory is obliged to answer this question, for however obligations arise, the theory should be effective in describing them. My sense is that obligations are created by social institutions and the individuals empowered (in the felicitous term of Jones and Sergot) to create obligations, but I think nothing here turns on that. Finally, if  $a$  in  $(\text{oblige}(a) \dots)$  is a state, what is it a state of? There is a state of being U.S. president and that

state currently attaches to George W. Bush. What does *a* attach to? Anything at all, so long as *a*'s unique identity can be maintained. Or so I suspect, but these are matters for subsequent ventures.

## References

- [1] A.R. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 67:100–3, 1958.
- [2] Donald Davidson. *Essays on Actions and Events*, chapter The Logical Form of Action Sentences, pages 105–148. Clarendon Press, Oxford University Press, Walton Street, Oxford OX2 6DP, United Kingdom, 1980. ISBN: 0-19-824637-4.
- [3] James Higginbotham, Fabio Pianesi, and Achille C. Varzi, editors. *Speaking of Events*. Oxford University Press, New York, NY, 2000. ISBN: 0-19-512811-7.
- [4] R. Hilpinen. Actions in deontic logic. In John-Jules Ch. Meyer and Roel J. Wieringa, editors, *Deontic Logic in Computer Science: Normative System Specification*, pages 95–100. John Wiley & Sons, New York, NY, 1993.
- [5] Steven Orla Kimbrough. Reasoning about the objects of attitudes and operators: Towards a disquotations theory for representation of propositional content. In *Proceedings of ICAIL '01, International Conference on Artificial Intelligence and Law*, 2001. Available at: <http://grace.wharton.upenn.edu/~sok/sokpapers/2000-1/icail/sok-icail01.pdf>.
- [6] Richard Larson and Gabriel Segal. *Knowledge of Meaning: An Introduction to Semantic Theory*. The MIT Press, Cambridge, Massachusetts, 1995. ISBN: 0-262-62100-2.
- [7] John-Jules Ch. Meyer and Roel J. Wieringa. Deontic logic: A concise overview. In John-Jules Ch. Meyer and Roel J. Wieringa, editors, *Deontic Logic in Computer Science: Normative System Specification*, pages 3–16. John Wiley & Sons, New York, NY, 1993.
- [8] Terence Parsons. *Events in the Semantics of English: A Study in Subatomic Semantics*. Current Studies in Linguistics. The MIT Press, Cambridge, MA, 1990. ISBN: 0-262-66093-8.

/\* \$Header: good-samaritan-disquotation.tex,v 1.2 2002/05/08 02:23:49 sok Exp \$ \*/