Absolute Obligations and Ordered Worlds
Author(s): Marvin Belzer and Barry Loewer
Source: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 72, No. 1 (Oct., 1993), pp. 47-70
Published by: Springer
Stable URL: https://www.jstor.org/stable/4320442
Accessed: 12-06-2019 15:40 UTC

MARVIN BELZER AND BARRY LOEWER

ABSOLUTE OBLIGATIONS AND ORDERED WORLDS

(Received in revised form 15 March 1993)

Normative theories typically entail judgments about the permissibility of actions. Some types of actions are said, in some theories, to be "*absolutely* impermissible" in the sense that there is no conceivable situation in which, all things considered, performance of an action of that type (for instance, torturing an innocent person) would be permissible. Other action types (for instance, stealing a loaf of bread) are not absolutely impermissible (even though they are "*prima facie* impermissible"), since there are possible although unusual situations in which it would be permissible, all things considered, for someone to perform an action of that type.

A deontic logic that would be useful in studying and applying substantive normative theories must represent the distinctions between absolute, prima facie, and "all things considered" permissions (and obligations). Now an ordering of possible worlds has been used in deontic logic to provide a semantical interpretation for expression of conditional obligatoriness and permissibility, for instance, by David Lewis [12]. We used a Lewis ranking of worlds as the core of a system of deontic logic called 3-D (cf. [3, 4, 15, 16]) which provides means for distinguishing between expressions of prima facie and all-things-considered normative judgments. In this article we consider expressions of absolute obligations. Our first goal is to show that there is an important class of normative systems whose content cannot accurately be represented by the Lewis rankings of 3-D, specifically, those normative systems that can generate more than one logically independent "absolute" obligation. Later we revise the semantics to handle these cases, but we provide reasons for thinking that any moral theory that generates independent absolute obligations must allow for the possibility of all-things-considered moral conflicts.

Philosophical Studies **72**: 47—70, 1993.

I

Lewis' semantics for expressions of obligations and permissions assumes that a normative system weakly orders a set of possible worlds according to their acceptability relative to that normative system. Let \$ be a function that assigns to normative systems n a weak ordering of worlds $\$_n$.[1] "It ought to be that B, given A," O(B/A), is true relative to n iff there are no A-worlds ranked in $\$_n$ or there is some A&B world ranked in $\$_n$ more highly than any A&-B world. We will say that O(B/A) is "vacuously true" relative to n iff there are no A-worlds ranked in $\$_n$. Also, let OA $=_{df}$ O(A/T), for tautology T. And as usual we define permissibility ("it is permissible that B, given A," P(B/A)) as −O(−B/A). O(B/A) is *defeasible* if there is some C such that −O(B/A&C). It is the consistency of O(B/A) and −O(B/A&C) that makes the Lewis semantics so useful in representing the content of normative systems. Consider for example a small normative system n existing in a certain situation with only these rules:

i. Smith should not tell the secret to Arabella if he tells Columbo.

ii. Smith should not tell the secret to Barbarella if he tells Columbo.

iii. Smith should tell Arabella (and it is not the case that he should not tell Arabella) if he tells both Columbo and Barbarella.

iv. Smith should tell Barbarella (and it is not the case that he should not tell Barbarella) if he tells both Columbo and Arabella.

These can be represented in Lewis as

i* O(−a/c)

ii* O(−b/c)

iii* O(a/b&c) & −O(−a/b&c)

iv* O(b/a&c) & −O(−b/a&c).

The rule O(−a/c) is defeasible because of the truth of −O(−a/b&c).[2]

To see how a ranking of worlds can be used to interpret these rules consider first the following eight types of worlds:

1	a	b	c
2	a	b	−c
3	a	−b	c
4	a	−b	−c
5	−a	b	c
6	−a	b	−c
7	−a	−b	c
8	−a	−b	−c

Any ranking of worlds (world-types) that makes true i*—iv* will have to rank either type 5 or type 7 more highly than both type 1 and type 3 (to make i* true), either 3 or 7 more highly than both 1 and 5 (for ii*), 1 more highly than 5 (for iii*), and 1 more highly than 3 (for iv*). There are a number of rankings that meet these conditions, for instance

$\$_n \qquad (2, 4, 6, 7, 8) < 1 < (3, 5)$

in which worlds of types 2, 4, 6, 7, 8 are ranked equally as most acceptable, worlds of type 1 next, and worlds of types 3 and 5 equally at the bottom.[3]

The settled facts of a particular situation may defeat O(B/A). Let "the settled facts" that hold in a particular situation (at time t in a world h) be represented by the set of worlds that are physically possible at t at h. If O(p/r) and −O(p/r&s) are true for n, and each accessible world at t, h is an r&s world (so that r&s is "settled" at t at h) then O(p/r) is defeated at t, h. Of course there is at t, h a **prima facie** obligation that p (due to O(p/r) and r being settled),[4] but expressions of prima facie obligation have limited value in guiding action. Yet we can define **all-things-considered (a.t.c.)** obligations relative to the class of settled facts at t, h by restricting the ranking $\$_n$ to worlds that are accessible at t, h: Let $\$_{n(t, h)}$ be just like $\$_n$ except that worlds that are inaccessible at t, h are not ranked in $\$_{n(t, h)}$. For example: if the ranking

$\$_n \qquad (2, 4, 6, 7, 8) < 1 < (3, 5)$

represents the content of the small normative system introduced above, and if c is settled as true at t, h but neither a, −a, b, nor −b are settled, then the restricted ranking $\$_{n(t,h)}$ for t, h is $\$_n$ with the world-types that make −c true removed, namely

$$\$_{n(t,h)} \quad 7 < 1 < (3, 5).$$

If at t* later than t, b also becomes settled then we have

$$\$_{n(t^*,h)} \quad 1 < 5.$$

The restricted rankings can be used to provide a truth condition for the expression of temporally-relative a.t.c. obligations, as follows. "It ought a.t.c. at time t to be that B, given A," $O_t(B/A)$, is true at h relative to n iff there is some A&B world ranked more highly in $\$_{n(t,h)}$ than any A&−B world. Expressions of a.t.c. obligations and permissions that are true in a particular situation *are* useful in guiding action in that situation. O_tp, that is $O_t(p/T)$, is true at h relative to n iff p is true at each of the most highly ranked worlds in $\$_{n(t,h)}$. Thus in the example O_t−a and O_{t*}a both are true at h.

This framework, which we call **3-D**, provides the foundation for a theory of normative reasoning where, typically, a.t.c. obligations and permissions are entailed by ideal obligations and permissions and settled facts (plus **ceteris paribus** statements). This theory of normative reasoning can be applied to any normative theory that has principles that can be represented accurately by means of a ranking of possible worlds.[5]

There are a number of other useful concepts that can be defined. Let us say first that the obligation expressed by O(B/A) is **absolute** ("@(B/A)") relative to a normative system n if, and only if, O(B/A&C) is true for n for all C for which A&B&C is consistent. The requirement that A&B&C be consistent is necessary here because relative to Lewis rankings there is no B that is non-vacuously obligatory given something that is inconsistent with B — that is, −O(B/−B) always holds if −B worlds are ranked in $\$_n$. (To see this suppose that some −B world k is ranked in $\$_n$, and suppose that k is a best −B world in $\$_n$. Since k is not a B world, −O(B/−B).) So we do not wish to define absolute obligations simply as "O(B/A) is absolute iff O(B/A&C) for all C" — for this definition entails that absolute obligations are possible only if no worlds inconsistent with A&B are ranked. It follows from our definition that OA is **absolute** ("@A") relative to n just in case there is no C such that both A&C is consistent and −O(A/C) is true relative to n. Absolute obligations expressed by @A are indefeasible relative to any situation that is compatible with A. For example, if the obligation to procure informed consent from the participants in a medical experiment is absolute, then there are no possible circumstances compatible with one's procuring informed consent in which it would be permissible "all things considered" not to do so.

Obligations also can be compared with respect to **relative weight**: O(B/A) has greater relative wight in n than O(D/C) just in case O(B/A&C&−(B ≡ D)) is non-vacuously true relative to n.[6] It follows that for unconditional OB and OD, OB has greater relative weight in n than OD just in case O(B/−(B ≡ D)). Any two non-vacuous obligations O(B/A) and O(D/C) will stand in determinate weight relationships because of the following principle:

(I) Not both O(B/A&C&−(B ≡ D)) and O(D/A&C&−(B ≡ D)) can be non-vacuously true.

One of the obligations will have greater weight than the other or they will have equal relative weight.[7]

II

As it is based on the Lewis ranking of worlds according to their acceptability, it may appear that the 3-D semantics is constructed for utilitarian consequentialists who readily compare worlds according to their possession of desired features and require actions insofar as they play a role in bringing about the *best* worlds accessible. This appearance, however, reflects a misunderstanding of the semantics. Worlds can be ordered according to the degree to which the norms of a non-consequentialist are satisfied in those worlds. For example, a world that is very bad according to a certain utilitarian nevertheless may be ranked as acceptable in the ordering of worlds that represents the content of a deontological theory; and worlds good for the utilitarian may be ranked as unacceptable in the deontological ranking. Lewis himself argued that his semantics is ethically neutral:

> a simplistic non-utilitarian might fancy an ordering on which the better of any two worlds is the one with fewer sins. (It is up to him to tell us how he divides the totality of sin into distinct units.) Under this ordering and my semantics, much is obligatory and little is permissible. Perhaps some of the worlds where Jesse robs the bank have sixteen sins, none have fewer, and some have more. Then what is obligatory, given that Jesse robs the bank, is that there be no seventeenth sin. No course of action with any extra sin is (even conditionally) permissible, no matter how much counterbalancing good there may be ([12], p. 86).

Lewis' example seems to show that a ranking of worlds can be used to represent the content of a non-consequentialist theory. It might be argued, however, that Lewis' "simplistic non-utilitarian" is in fact a consequentialist insofar as he maximizes sinlessness. Suppose for instance that there is a simplistic non-utilitarian who prohibits only the eating of ice cream on Sunday. On Sunday you find yourself in a "Scheffleresque" situation: it is settled that if you eat ice cream five others will *not*, and if you do not eat ice cream then the five others *will* (cf. [20] for discussion of this type of case). If the proper goal is to maximize sinlessness, then you should eat the ice cream — after all, one violation is better than five. But of course a non-consequentialist may prohibit your eating ice cream even in this situation: eating ice cream even may be *absolutely* prohibited.

Now these points do not show that the Lewis ranking can be used for no non-consequentialist theories. A *very* simple non-consequentialist theory may require simply that James Brown eat no ice cream on July 1, 1989. The content of this normative "theory" (and others much less simple) can be presented satisfactorily by a Lewis ranking.

However we now would like to show that there is an important class of normative theories, those that endorse logically independent absolute obligations or prohibitions, whose content cannot be represented accurately by means of a Lewis ranking of worlds.

Consider a normative system n that both prohibits absolutely your telling a certain secret to Penrose and also prohibits absolutely your telling it to Sally. Under no circumstances is it permissible for you to tell the secret to either Penrose or Sally. We can represent the absolute prohibitions as $@-p$ and $@-s$. This assumption that there are two absolute prohibitions $@-p$ and $@-s$, for logically independent $-p$ and $-s$, leads to a contradiction if there are $-(p \equiv s)$ worlds ranked in $\$_n$. (By "logically independent" we mean the following. A and B are logically independent just in case A entails neither B nor −B and B entails neither A nor −A.) The proof is as follows. Let Mp say "p is logically possible." By our definition of "absolute" we know

$$@-p \text{ iff } O-p \;\&\; (q)(M(-p\&q) \rightarrow O(-p/q)).$$

Now $-p\&-(p\equiv s)$ is consistent, that is, $M(-p\&-(p\equiv s))$, so $O(-p/-(p\equiv s))$. Similarly

$$@-s \text{ iff } O-s \;\&\; (q)(M(-s\&q) \rightarrow O(-s/q)).$$

Again $-s\&-(p\equiv s)$ is consistent, i.e. $M(-s\&-(p\equiv s))$, so $O(-s/-(p\equiv s))$. Suppose that there are $-(p\equiv s)$ worlds ranked; then both $O(-p/-(p\equiv s))$ and $O(-s/-(p\equiv s))$ are non-vacuously true. But this contradicts principle (I) which entails that not both $O(-s/-(p\equiv s))$ and $O(-p/-(p\equiv s))$ can be non-vacuously true. Therefore if A and B are logically independent and $\$_n$ ranks $-(A\equiv B)$ worlds, then $\$_n$ permits the truth of **at most one** of @A and @B. So $\$_n$ cannot reflect accurately the content of a normative system that absolutely prohibits A and also absolutely prohibits B.[8]

The argument does not go through if $-(p\equiv s)$ worlds are inevaluable in n, that is, if no such worlds are ranked in $\$_n$. It might be objected that any system that absolutely prohibits p and s cannot coherently rank $-(p\equiv s)$ worlds. If this claim were correct then our argument against the Lewis semantics would fail. But the claim is not correct since n might also require your telling Columbo if you tell the secret to either Sally or Penrose but not both, that is $O(c/-(p\equiv s))$ is non-vacuously true. This additional obligation is consistent intuitively with @−p and @−s, but of course $O(c/-(p\equiv s))$ can be non-vacuously true only if $-(p\equiv s)$ worlds are ranked in $\$_n$.[9] The objection confuses two distinct notions — absolute obligations and what might be called the "existence conditions" for a normative system. Let A be an existence condition for a normative system n just in case the "existence" (or instantiation) of n at a time t in a world k requires that −A not be settled at t, k. If −A is settled at t, k then n cannot exist. Thus it is plausible to say that if A is an existence condition for n, $\$_n$ cannot coherently rank any −A worlds.[10]

On the other hand it might be suggested that the argument shows only that our definition of **absolute** is defective rather than that the

Lewis ranking is not suitable for the representation of certain types of normative systems. However our conception of absolute obligation and prohibition is not central to the problem that has been exhibited. The real shortcoming of the Lewis semantics is principle (I). A normative system may wish to incorporate both

$$O(-p/-(p \equiv s))$$

and

$$O(-s/-(p \equiv s))$$

even in cases in which neither p nor s are absolutely prohibited (that is, even when there are q and r compatible with −p and −s, respectively, such that $-O(-p/q)$ and $-O(-s/r)$). Therefore our problem is independent of our definition of absolute prohibitions, although the problem is illustrated most strikingly by considering absolute prohibitions as we have defined them. The semantics forces a comparison of relative weight between any two non-vacuous obligations O−p and O−s; if neither has greater weight than the other then they are regarded as having "equal weight" in a framework in which "equality of weight" means

$$-O(-p/-(p \equiv s)) \;\&\; -O(-s/-(p \equiv s)),$$

that is

$$P(p/-(p \equiv s)) \;\&\; P(s/-(p \equiv s)).$$

Yet this means that **both** prohibitions are removed in any case in which both (a) neither O−p nor O−s has explicitly been given greater relative weight than the other by virtue of the truth of $O(-p/p \equiv s)$ or $O(-s/p \equiv s)$ respectively, and (b) it is settled that one but not both of the prohibitions is to be satisfied.

The Lewis ranking semantics is prejudiced in favor of consequentialist theories insofar as *typically* they do endorse only one independent absolute obligation (e.g. to maximize happiness) whereas non-consequentialist theories often endorse independent absolute obligations. Our argument therefore confirms Foot's suspicion that a non-consequentialist may grant the consequentialist too much — and thereby be taken in by a "conjuring trick" — "in accepting the idea that there *are* better and worse states of affairs in the sense that consequentialism requires" ([11], p. 199). In a Lewis ranking worlds are ranked according to acceptability, but some normative theories are not coherent if we insist that the content of a normative theory be given by such a ranking of worlds. Consider for instance this passage from Feldman's discussion moral integrity in hard cases:

So long as people find certain act types morally repugnant, every normative theory runs the risk of violating someone's moral integrity. Suppose someone finds acts of type M (murder, for example) to be morally repugnant. Suppose this person now finds himself in a horrible situation in which he must either murder A or murder B. *Any normative theory that generates any prescription for this situation violates the moral integrity of this poor person* ([9], p. 67, emphasis added).

The emphasized sentence ignores the possibility that a normative theory will contain both (a) $O(-A/A \vee B)$ and (b) $O(-B/A \vee B)$, in which case the theory generates a prescription for the poor fellow (indeed two of them) and yet does not violate the person's moral integrity (at least not by requiring him to do something absolutely impermissible). By absolutely prohibiting both A and B, a system can generate prescriptions here without violating moral integrity. Of course it is quite true that both (a) and (b) cannot be true relative to a Lewis ranking that ranks some $(A \vee B)$-worlds (and Feldman's work like our 3-D is based on this sort of semantics) but that is a formal objection to a normative theory that does contain both (a) and (b) only if the Lewis ranking of worlds is the only sort of semantics, or the best sort of semantics, that can be constructed for expression of obligations and permissions.

As for the fact that the theory will generate two prescriptions in Feldman's example — it will both require $-A$ and require $-B$ even though it is settled that $A \vee B$: this is "one too many" because of course if $A \vee B$ is settled, then at least one of the absolute prohibitions $@-A$ and $@-B$ will have to be violated. The system is not as helpful as might be wished in guiding one's action, because it leaves one in a quandary, but perhaps it is a mistake to expect a moral system to provide unique and complete guidance in any situation in which there are competing grounds of obligation (cf. Donagan [8], pp. 307—308). A system that does not create a conflict in this situation may require that one not murder both A and B, given $A \vee B$, even if it does not provide complete guidance (since the question whether actually to murder A or to murder B may not be a moral question, cf. [8], p. 307; Feldman also makes this point, [9], p. 201). Therefore in the example such a theory also could generate a relevant prescription, namely O(−(A&B)/A ∨ B), without violating moral integrity.

Even if we have found that the Lewis rankings favor consequentialist theories, it nonetheless is worth noting that a consequentialist theory could employ two independent (incommensurable) absolute standards in which case the Lewis ranking would not even be applicable to it. And, on the other hand, Lewis rankings are applicable to non-consequentialist theories that do not endorse independent absolute obligations or prohibitions. Notice also that even theories with only one "overriding principle" may under certain conditions endorse more than one independent absolute obligation — e.g. a principle making murder absolutely impermissible can ground the independent absolute obligations expressed by @−p and @−s in Feldman's example.

III

A simple way of revising the 3-D semantics to encompass systems that do not conform to Lewis is to introduce the "superposed" worlds of Rescher and Brandom [18]. A superposed world k is a construction from two worlds h and j such that a proposition A is true at k iff A is true at h or A is true at j. Both A and −A may be true at a superposed world k even though no classical laws of logic are violated at k. Suppose for instance that k is a superposition of h and j, where A is true at h and −A is true at j. Then both A and −A are true at k, but since A&−A is true at neither h nor j, A&−A also is not true at k. If superposed worlds in addition to normal worlds are eligible to be ranked in $\$_n$, then it can turn out that each best A world is a superposed world at which both B and −B are true, in which case both O(B/A) and O(−B/A) are true relative to n. Revise the truth condition for O(B/A) so that it is true non-vacuously iff all most highly ranked A worlds are B worlds. (This is equivalent to the original truth condition if only normal worlds are ranked, but not so given superposed worlds.) Also our earlier definition of "relative weight" needs supplementation, as follows:

> O(B/A) has greater relative weight than O(D/C) in a normative system n just in case O(B/A&C&−(B≡D)) is true relative to n *and O(D/A&C&−(B≡ D)) is not true relative to n.*

The emphasized clause is now needed because it no longer is guaranteed by the first clause on the right. If both O(B/A&C&−(B≡D)) and O(D/A&C&−(B≡D)) are true relative to n then we can say that O(B/A) and O(D/C) are **incommensurable** in n.

Principle (I) no longer is valid, for it may be that each best A&C&−(B≡D) world in the ideal field for n is a superposed world at which B and D are true, in which case both

O(B/A&C&−(B≡D))

and

O(D/A&D&−(B≡D))

are true.[11] Hence a deontological system like the one we discussed above to show that the Lewis ranking is not ethically neutral may consistently contain both

(c) O(−p/−(p≡s))

and

(d) O(−s/−(p≡s));

and both @−p and @−s hold even though −p and −s are logically independent.

Much of the 3-D structure remains intact even after superposed worlds are introduced. We need to expand the definition of "accessibility" if we are to permit the accessibility of superposed worlds at t in h. Let k be **s-accessible** at t in h just in case either (i) k is a normal world and accessible at t in h or (ii) k is a superposed world and such that for each proposition A true at k, there is some normal A-world accessible at t in h. The restricted ranking $\$_{n(t,h)}$ is like $\$_n$ except that worlds that are not s-accessible at t, h are excluded from $\$_{n(t,h)}$. This means that a superposed world k that is ranked in $\$_n$ will also be ranked in $\$_{n(t,h)}$ only if k makes true no proposition A which is such that $-A$ is settled at t, h. Thus for example if k is a superposition of k_1 and k_2 where A is true at k_1 and $-A$ is true at k_2 then k is s-accessible at t, h only if neither A nor $-A$ is settled at t, h.

The new semantics countenances the possibility of a.t.c. conflicts; that is it may be the case that both O_tA and O_t-A are true for n at h. (Suppose that each most highly ranked world in $\$_{n(t,h)}$ is a superposed world that makes true both A and $-A$.) There is no incoherence since the principle of "agglomeration"

$$O_tA \& O_tB \rightarrow O_t(A\&B)$$

no longer is valid in the new logic,[12] even though "ought implies can" remains valid (in the sense that O_tA can be true at t at h relative to n only if there is some normal A world that is accessible at t, h).[13] Donagan has argued as follows on behalf of the agglomeration principle for moral systems (and since he also accepts "ought implies can" his argument is directed against the possibility of a.t.c. conflicts):

> autonomous moral agents will reject a moral system as ill constructed if there are situations to which its precepts apply, but in which their agglomeration would be invalid ([8], p. 300).

His argument relies upon the assumption that the agglomeration principle fails only in "systems of commands by appropriate authorities" ([8], p. 298). Now it *may* be the case that any proposed moral system that violates agglomeration would for that reason be deficient and unacceptable, like systems of commands given by unreasonable authorities.

We would conjecture, however, that any semantics that allows for the consistency of (c) and (d) also will have to countenance the possibility of a.t.c. conflicts, simply because of the possibility of situations in which $-(p \equiv s)$ is settled and nothing else is relevant. Some deontologists like Kant and Donagan hold that a.t.c. conflicts are not possible, yet they also endorse principles that can give rise to several independent absolute prohibitions. If the conjecture is true, then (c) and (d) will be inconsistent in systems that do not permit a.t.c. conflicts, which means that there cannot exist in such systems two distinct absolute prohibitions $@-p$ and $@-s$ where $-p$ and $-s$ are logically independent. (It follows that for all we have shown in this paper, deontological systems that do not permit a.t.c. conflicts can be represented accurately by means of the Lewis ranking even without superposed worlds.) If we are correct, then one could not coherently endorse distinct absolute prohibitions while also rejecting the possibility of a.t.c. conflicts. Of course one might have reasons for thinking that they never in fact occur even though they are possible.

Non-consequentialists insist that even though right and wrong is not determined by the relative goodness of states of affairs according to some independent standard, nevertheless such teleological considerations play an important role in morality. Foot for instance argues that the idea of "maximum welfare" appears "*within* morality as the end of one of its virtues [benevolence]" even though it does not stand "*outside* morality as its foundation and arbiter" ([11], p. 206). Donagan, while claiming that the notion of "maximizing ends in themselves" is nonsensical (because only persons are ends in themselves), nonetheless holds that

> producing more rather than less goods for ends in themselves is a secondary rational end, that is, one it would be self-contradictory not to have if you recognize that there are ends in themselves ("The relation of moral theory to moral judgments: a Kantian view," mimeo, p. 7).

We will consider briefly how a moral teleology can be embedded into a deontological moral theory within the **3-D** structure (whether or not superposed worlds are ranked) and indicated how this embedding may be useful in reasoning in a deontological system.[14]

Suppose given an ordering X of worlds according to the welfare of persons in those worlds. This ordering represents the content of a substantive theory of welfare which (we will suppose) is part of a general moral theory. Also suppose given a possibly quite different ordering Y of worlds according to their acceptability relative to the principles of the moral theory: Y is the ideal field of moral permissibility for the theory. Now of course some worlds in Y may be ranked equally; let the set of worlds ranked equally with a world k be called the "level" of k. The worlds in the level of k in Y need not be ranked equally in the teleological ordering X. On the contrary, some may be vastly more acceptable than others according to X. At each level Y we can rank the worlds on that level according to their ranking in the teleological ranking X. In a situation in which no worlds better than some of those in the level of k are accessible, we know that any accessible world in that level is *a.t.c. permissible* but all the same some of the accessible worlds in that level may be regarded as *better* than others.

Suppose that a deontological system were to insist as in Feldman's example discussed above that any murder is absolutely forbidden. The system refuses to measure the relative acceptability of the murder of A and the murder of B. As above, we have simply

$$O(-A/A \lor B)$$

and

$$O(-B/A \lor B).$$

Now if each of the best A $\lor$ B worlds is both a $-$A world and a $-$B world, then each such world must be a superposed world. And if as in Feldman's example one is in a situation in which A $\lor$ B is settled and nothing else is relevant, then both A and B are forbidden in that situation, relative to the deontological system we are discussing, for each most acceptable world in the range of accessibility will be superposed world at which both $-$A and $-$B are true.[15]

Even though both A and B are a.t.c. forbidden in this situation, the moral theory nevertheless may provide additional guidance because of the embedded teleological ranking X of worlds. For it may be, for instance, that some of the most acceptable accessible worlds at which $-$A, B and $-$B (but *not* A) are true are ranked more highly in X than any of the most acceptable accessible worlds at which $-$B, A and $-$A (but not $-$B) are true. This is grounds for saying that even though A and B are both absolutely wrong, nevertheless it is morally preferable to choose B over A in this case. It might be said that A is a *graver* wrong than B. Cf. Donagan:

> although wrongness, or moral impermissibility, does not have degrees, impermissible wrongs are more or less grave . . . It is absolutely impermissible either to murder or to steal; but although murder is no more a wrong than stealing, it is a graver wrong ([7], p. 152).

The teleological ranking is not *always* necessary for the application of Donagan's "principle of the least evil" ("when you must choose between evils, choose the least," [7], p. 152) because if both O$-$p and O$-$s are true (so that both p and s are "evils") and if O$-$p has greater relative weight than O−s (so that if nothing else is relevant except that one of the two will be violated one should violate O−s), then the system of rules even without the teleological ranking provides guidance as to choosing the least evil. But its embedding can provide additional guidance even when two absolute prohibitions (with equal or incommensurable relative weight) are involved.

IV

While the notion of superposed worlds provides a technical fix for the representation of independent absolute obligations in 3-D it nonetheless is not an entirely satisfying idea. Formerly we pruned the Lewis-ranking $\$_n$ with the shears of settled fact to get the action-guiding ranking $\$_{n(t, h)}$: the set of worlds most highly ranked in $\$_{n(t, h)}$ was the set of worlds that it would be permissible, all things considered at t in h, *to bring about*. This set of course was a subset of the worlds accessible at t, h. But now, with superposed worlds being ranked in $\$_n$, it may be that the set of accessible worlds at t, h and the set of a.t.c. permissible worlds at t, h are disjoint (as is the case if each most highly ranked world in $\$_{n(t, h)}$ is superposed) — in which case the set of a.t.c. permissible worlds fails to provide guidance as to which worlds may be brought about. Of course this may be just what one should expect if a.t.c. conflicts O_tA and $O_t{-}A$ are possible. But if so the attractive features of the 3-D framework have been saved only superficially for systems whose content cannot be represented by a ranking of (normal) worlds.

In any case the Lewis-ranking based semantics faces an additional difficulty whether or not superposed worlds are ranked. If O(B/A) is interpreted as expressing conditional obligations (or, more generally conditional "oughts"), then the settled truth of A would provide a reason in support of the claim that B is a.t.c. obligatory (or, that B ought a.t.c. to be the case). This at least is a plausible construal of O(B/A).[16] However in the ranking based semantics O(A/A) is true whenever any A worlds are ranked — even if −A is absolutely prohibited, that is, @−A. We showed above that a normative system coherently may rank A worlds even though @−A, but it seems odd to hold that the settled truth of A would provide a reason in support of the claim that A is a.t.c. obligatory (even if its settled truth would provide a non-vacuous reason for other things B, e.g. repentence, punishment, an apology).

We have constructed a new deontic logic "General Deontic Logic" (GDL) whose language is that of 3-D and whose semantics has many of the virtues of 3-D but fewer of its vices, including those vices due to its employment of the Lewis ranking of worlds. To interpret O(B/A) let F be a function from normative systems and propositions (sets of worlds) to sets of propositions. Let [p] be the set of worlds at which p is true. $F_n([p])$ is the set of propositions required by p according to n. We place no conditions on F. The truth condition for O(Q/P) is as follows:

$$O(Q/P) \text{ is true relative to n iff } [Q] \,\varepsilon\, F_n([P]).$$

The logic of O(−/−) contains only the following two rules concerning substitution:

$$\text{RCOEA.}\quad \frac{P \equiv P'}{O(Q/P) \equiv O(Q/P')}$$

$$\text{RCOEC.}\quad \frac{Q \equiv Q'}{O(Q/P) \equiv O(Q'/P).}$$ [17]

This is an austere conditional deontic logic validating none of the following schemas:

O(T/T)
O(T/P) (T-requiring)
O(P/P) (idempotence)
−O(−P/P)
O(Q/P) & O(P/T) → O(Q/T) (deontic detachment)
O(Q/P) → −O(−Q/P) (no conflicts)
−O(P/F)
O(Q/P) & O(R/P) → O(Q&R/P) (agglomeration)
O(Q&R/P) → O(Q/P) & O(R/P) (decomposition)
O(Q/P) & O(Q/R) → O(Q/P ∨ R)
O(Q/P ∨ R) → O(Q/P) ∨ O(Q/R).
O(Q/P) → O(Q/P&R) (augmentation)
O(Q/P) & P → OQ (O-mp)

$O(Q/P) \& -O(-R/P) \rightarrow O(Q/P\&R)$ (von Wright)
$O(Q/P) \& -O(Q/-Q \vee P) \rightarrow O(-P/-Q)$ (transposition*)
$O(Q/P) \& O(R/Q) \& -O(-P/T) \rightarrow O(R/P)$ (restricted chaining.)

The logic is even more "minimal than Chellas' "minimal conditional deontic logic" CD which validates

$$O(Q\&R/P) \rightarrow O(Q/P) \& O(R/P) \quad \text{(decomposition)}$$

and the following rule not valid here:

$$\text{RCOM. } \frac{B \rightarrow B'}{O(B/A) \rightarrow O(B'/A)}$$

(cf.) Chellas [6]). RCOM is valid given the condition on F_n that

(cm) if $x \cap y \varepsilon F_n(z)$ then $x \varepsilon F_n(z)$ and $y \varepsilon F_n(z)$.

Because CD has the rule RCOM it is subject to the "gentle murder" paradox (cf. [10, 16]): let m represent "Smith murders Jones" and g represent "Smith murders Jones gently." In a normative system in which O(g/m) is true, where g implies m, RCOM also gives O(m/m). But m need not be regarded as requiring (or being a reason for) m, even though g requires m.

To interpret expressions of a.t.c. obligation, first let a function f assign sets of propositions f_k to world-time pairs $\langle t, h\rangle = k$. We will suppose f_k is closed under the usual Boolean operations. Let $S_t p$ say that p is "settled" at t:

$S_t p$ is true at h iff $p \varepsilon f_k$.

If p is settled at t, then given O(q/p) there is some reason for q; but not necessarily a "conclusive" reason, since possibly −O(q/p&r) for some r that also is settled at t. Let $O_t p$ say as before that there is a conclusive reason for p in situation k, i.e. "all things considered in k it ought at t to be that p."

$O_t p$ is true at h relative to n iff there is some q such that $[q] \varepsilon f_k$ and $[p] \varepsilon F_n([q])$ and there is no $x \varepsilon f_k$ such that not $[p] \varepsilon F_n([q] \cap x)$.

Notice that O(p/q) and $S_t q$ do not entail $O_t p$; we need to know in addition that there is no r such that $S_t r$ for which −O(p/q&r). In other words, to detach $O_t p$ from O(p/q) and $S_t q$ we need to know that q is sufficient information in the circumstances at t, h to determine conclusively the normative status of p (relative to n). Let $R_t(p, q)$ say as much.

> $R_t(p, q)$ is true at h relative to n iff there is no $x \varepsilon f_k$ such that not $[p] \varepsilon F_n([q] \cap x)$.

The following detachment schema is valid:

$$\text{(key) } O(p/q) \;\&\; S_t q \;\&\; R_t(p, q) \rightarrow O_t p.$$

The logic of O_t also is austere, containing only the following substitution rule

$$(\text{RCO}_t\text{E}) \quad \frac{A \equiv A'}{O_t A \equiv O_t A'}$$

It is not closed under implications since the following rule fails:

$$(\text{RCO}_t\text{M}) \quad \frac{B \rightarrow B'}{O_t B \rightarrow O_t B'}$$

and none of the following schemas is valid

$O_t(P \rightarrow Q) \;\&\; O_t P \rightarrow O_t Q$ (O_t-detachment)
$O_t Q \rightarrow -O_t-Q$ (no O_t-conflicts)
$-O_t F$
$O_t Q \;\&\; O_t P \rightarrow O_t(Q\&P)$ (O_t-agglomeration)
$O_t(Q\&P) \rightarrow O_t Q \;\&\; O_t P$ (O_t-decomposition)
$O_t(P \rightarrow Q) \;\&\; S_t P \rightarrow O_t P$ (t-mp)
$O_t P \;\&\; S_t(P \rightarrow Q) \rightarrow O_t Q$ (t-necessary conditions)

Since RCO_tM fails the logic is immune to the "gentle murder" paradox and other versions of the good samaritan paradox [16]. Meanwhile the logic can handle satisfactorily "contrary to duty" paradoxes like Chisholm's paradox (cf. [15]) which has a structure similar to the "secrets" story given above with rule i*—iv*

i* O(−a/c)

ii* O(−b/c)

iii* $O(a/b\&c) \& -O(-a/b\&c)$

iv* $O(b/a\&c) \& -O(-b/a\&c)$.

Given S_tc and $R_t(-a, c)$ we can detach O_t-a by means of the "key" schema. Later at t* when b also is settled, $R_{t*}(-a, c)$ no longer is true. But $R_{t*}(a, b \& c)$ is true and we can detach $O_{t*}a$ using iii′.

In GDL we can also characterize a notion of absolute obligation as follows:

$$@q \text{ iff for all r for which } F_n(r) \text{ is defined, } O(q/r).$$

Nothing in GDL prevents there being a number of independent absolute obligations. Notice the difference between @q (it ought absolutely to be that q) and O_tq (it ought all things considered at t to be that q). The former is situation independent and says that the obligation that q is not overridden in any possible situation while the latter is situation dependent and says that in situation t the obligation that q is not overridden.

As noted GDL is rather weak, in fact so weak that there are important types of normative reasoning that cannot be represented. For instance, in means-end reasoning of the simplest type one knows that O_tp and in the circumstances it is settled that if p then q; so (it often seems) one should infer O_tq. E.g. you ought now, all things considered, to feed the cats tonight, and it is settled now that you will feed the cats tonight only if you go to the store by 5 p.m. to buy cat food; so you ought now, all things considered, go to the store by 5 p.m. to buy cat food. Of course if p logically implies q then it will be settled in any situation that if p then q. The feature of GDL that enables it to escape the good samaritan paradoxes also limits its use in formalizing normative reasoning.

One solution is simply to keep track of two (perhaps more) different all-things-considered "O_t"s, the first being like the one introduced above, the second being closed under "settled" implication. The good samaritan paradoxes still can be defused because (e.g. in the gentle murder) O_tg but not O_tm can be derived in the first sense of O_t even though both are derivable in the second sense. Focussing in this manner on the a.t.c. ought, however, we do not validate RCOM and yet one might think that there also is an important role for normative reasoning that is dependent on RCOM as well. E.g. O(you buy cat food/cat is hungry) certainly seems to entail O(you buy something/the cat is hungry). As we show elsewhere [5] GDL can be strengthened in stages so as to obtain progressively stronger systems of deontic logic with the strongest being 3-D.[18]

APPENDIX

In this appendix we will attempt to refute the claim that the rules in the set {i—iv} from section I can be reconstructed into a set of indefeasible rules. The key to the reconstruction, it might be said, is to specify in the antecedents of each of the allegedly defeasible principles the negations of the conditions that might defeat those principles. That is, the allegedly defeasible principles i and ii can be replaced by the following indefeasible rules:

i#. Smith should not tell the secret to Arabella if he tells Columbo *and does not tell Barbarella.*

ii#. Smith should not tell the secret to Barbarella if he tells Columbo *and does not tell Arabella.*

The set {i#, ii#, iii, iv} contains no defeasible principles. Given this reconstruction it also might be thought that defeasibility always can be eliminated unless it is due to system indeterminacy or incompleteness.

This proposal is unacceptable because the reconstructed set {i#, ii#, iii, iv} actually contains less information than the set {i—iv}. Defeasibility has been eliminated at the cost of reducing information content. To see this suppose that Smith tells the secret to Columbo and also to both Arabella and Barbarella. Relative to the original norms of the story it is obvious that Smith has failed to satisfy his obligations (for he was obligated *not* to tell either Arabella or Barbarella if he were to tell Columbo). Relative to {i—iv} this failure can be explained simply as a violation of i and ii. *But none of the rules in* {i#, *ii#, iii, iv*} *has been violated*! This set requires only that *a* and *b* have the same truth value, if *c* is true — and *that* has been satisfied since Smith told all three. The original norms, however, require not only the same truth values, given c, but that the truth values be *false*! Therefore the reconstructed set omits potentially useful information from the normative system. This information is useful not only in evaluating actions relative to the system in order to praise, blame, reward, and punish, but also is useful in situations in which factual information is incomplete. If for instance Smith for some reason does not know whether or not he has told the secret to Barbarella but does know that he has told Columbo, then {i—iv} favors not telling Arabella but {i#, ii#, iii, iv} does not. So the original set provides guidance in situations in which the reconstructed set does not, another consequence of the reduced information content of the reconstructed set. Since the original set contains all of the information in the norms of the story whereas the reconstructed set does not, the objection should be regarded as failing to show that defeasible principles can be eliminated in the suggested manner.

Our response to the objection does not show that there are no situations in which allegedly defeasible principles can be reconstructed safely into indefeasible rules. But given the necessity of genuinely defeasible principles in certain contexts, the "reconstruction" policy cannot be applied universally. And even in contexts in which the reconstruction policy does not entail loss of information, the use of defeasible principles can significantly simplify the representation of the content of the normative system.

NOTES

[1] We let reference to worlds be implicit assuming for simplicity that the content of n is the same at all worlds.

[2] Do we really need defeasible rules like i′ to represent i? It is natural to suggest that it would not be difficult to reconstruct the set {i—iv} into a set of defeasible rules by specifying in the antecedents of each of the allegedly defeasible principles the negations of the conditions that might defeat those principles. As we show in the Appendix to this paper, however, this suggestion cannot be made to work.

[3] The ranking $\S_n$ ranks each world as highly as is compatible with satisfaction of the rules i—iv. Although other rankings also satisfy i—iv, for instance

$$\S_n 2 \qquad (4, 6, 7, 8) < 1 < (2, 3, 5),$$

it can be argued that $\S_n$ is preferable to any such alternative because the alternative makes true an O statement not explicitly required by the rule set that $\S_n$ does not make true. In this example $\S_n 2$ makes true O(−a/b&−c) while $\S_n$ does not and the rules i—iv do not seem to require this either.

[4] This notion can be defined as follows: "it ought **prima facie** to be that p" is true at t

at h relative to n iff there is some A such that A is settled at t at h and O(p/A) is true relative to n.

[5] It also provides a framework in terms of which it is possible to formalize the relationship between moral theories (sets of principles), settled facts, and individual moral judgments (expressions of a.t.c. obligations and permissions), cf. [3, 4, 15, 16] for discussions of normative reasoning in **3-D**. The original version of **3-D** in [15] was subject to objections pointed out to us by Tom Blackson and Charles Donahue which subsequent versions (cf. [4]) escape.

[6] For discussion of this conception of relative weight, cf. [1, 2].

[7] Proof of (I). Suppose O(B/A&C&−(B ≡ D)) is true non-vacuously, so there is some A&C&−(B ≡ D) & (B&−D) world k that is ranked more highly than any A&C& −(B ≡ D) & (−B&D) world. Now suppose for reductio that O(D/A&C&−(B ≡ D)) also were true. Then some A&C&−(B ≡ D) & (−B&D) world would be ranked more highly than any A&C&−(B ≡ D) & (B&−D) world, which contradicts the first supposition. So if O(B/A&C&−(B ≡ D)) then −O(D/A&C&−(B ≡ D)). QED. Geoffrey Sayre-McCord [19] pointed out that on Lewis' truth condition both of the O statements in (I) can be vacuously true.

[8] We *can* get @(−p&−s) in our example but the "decomposition" principle

$$@(-p\&-s) \rightarrow @-p$$

fails. If p1 is the proposition corresponding to the set of worlds relatively most acceptable in a Lewis ranking (with the Limit Assumption), then @p1 is true. If p2 is the set of worlds ranked more highly than any except those in p1, then @(p1 ∨ p2) also is true. If p3 is the set of worlds ranked more highly than any except those in p1 or p2, then @(p1 ∨ p2 ∨ p3) also is true; and so forth. @p1 is the only absolute obligation relative to a Lewis ranking that is not entailed by any other absolute obligation.

[9] Notice that a normative system coherently can make provision for the violation of absolute obligations even in the Lewis ranking semantics. To see this suppose the only rules of the normative systems are @−p and O(c/p) for independent p and c. There are 4 world types to consider

1	p c
2	p −c
3	−p c
4	−p −c

The ranking

$$(3, 4) < 1 < 2$$

makes both @−p and O(c/p) true non-vacuously.

[10] In which case O(C/−A) is vacuously true for all C, given Lewis' truth condition. Since −A makes n impossible it might be thought better to make O(C/−A) vacuously false — as can be done by slightly revising the truth condition (cf. Lewis [13]). If God's existence is an existence condition for n then it would be true to say that if God is dead everything is permitted (according to n).

[11] To see this suppose that there is one most acceptable A&C&−(B ≡ D) world k, and that k is superposition of worlds h and j such that B and −D are true at h and −B and D are true at j. Then −(B ≡ D) is true at k because true at both h and j, and each best −(B ≡ D) world is a B world and also a D world. Of course in this case we also have −B and −D, as well as B ≡ D, true at each best A&C&−(B ≡ D) world. To avoid this

result one simply needs more worlds: suppose that there are three most acceptable A&C&−(B ≡ D) worlds h, j, k where h is a superposition of h_1 and h_2, j a superposition of j_1 and j_2, and k a superposition of k_1 and k_2, where h_1 makes true A, C, −B, D; h_2: A, C, B, −D; j_1: A, C, B, D; j_2: A, C, −B, D; k_1: A, C, B, −D; and k_2: A, C, B, D. Now −B is not true at k, −D is not true at j, and B ≡ D is not true at h. But A&C&−(B ≡ D) is true at each of the three superposed worlds, as are both B and D.

[12] The principle O(B/A) → P(B/A) also fails now but P(B/A) ≡ −O(−B/A) remains valid.

[13] It is easily shown that agglomeration and "ought implies can" (as interpreted here) entail the absence of a.t.c. conflicts on the assumption (which we maintain) that A&−A is true at no accessible world.

[14] These suggestions are similar to proposals made by Sayre-McCord [19] and he should be regarded as a co-author of this section of the paper.

[15] Each such world must also be an A world or a B world in order to guarantee that each such world is an A ∨ B world.

[16] We are grateful to Walter Sinnott-Armstrong for emphasizing this point.

[17] To see that the logic has these rules it is sufficient to observe that if P ≡ Q then for any model [P] = [Q]; which means that if x = y, then both $z \varepsilon F(x)$ iff $z \varepsilon F(y)$, and $x \varepsilon F(z)$ iff $y \varepsilon F(z)$.

[18] A shorter version of this paper was presented on March 26, 1987, at the American Philosophical Association, Pacific Division, in San Francisco. We are grateful to Geoffrey Sayre-McCord for his comments.

BIBLIOGRAPHY

1. Belzer, M.: 1985, "Normative Kinematics (I): a Solution to a Problem about Permission," *Law and Philosophy* 4, 257–287.
2. ——: 1985, "Normative Kinematics (II): the Introduction of Imperatives," *Law and Philosophy* 4, 377–403.
3. ——: 1986, "Reasoning with Defeasible Principles," *Synthese* 66, 135–158.
4. ——: 1986, "A Logic of Deliberation," *Proc. Fifth National Conference on Artificial Intelligence*, vol. 1, 38–43.
5. Belzer, M. and Loewer, B.: 1993, "Logics of Defeasible Normative Reasoning," Society for Exact Philosophy, Toronto, May 16, 1993.
6. Chellas, B.: 1980, *Modal Logic*, Cambridge University Press, Cambridge.
7. Donagan, A.: 1977, *The Theory of Morality*, The University of Chicago Press, Chicago.
8. ——: 1984, "Consistency in Rationalist Moral Systems," *Journal of Philosophy* LXXXI, 291–309.
9. Feldman, F.: 1986, *Doing the Best We Can*, D. Reidel, Dordrecht.
10. Forrester, J.: 1984, "Gentle Murder, or the Adverbial Samaritan," *Journal of Philosophy* LXXXXI.
11. Foot, P.: 1985, "Utilitarianism and the Virtues," *Mind* 94, 196–209.
12. Lewis, D.: 1973, *Counterfactuals*, Basil Blackwell, Oxford.
13. ——: 1974, "Semantic Analyses for Dyadic Deontic Logic," in S. Stenlund (ed.), *Logical Theory and Semantic Analysis: Essays Dedicated to Stig Kanger*, Reidel, Dordrecht, pp. 1–14.
14. ——: 1978, "Reply to McMichael," *Analysis* 38, 85–86.
15. Loewer, B. and Belzer, M.: 1983, "Dyadic Deontic Detachment," *Synthese* 54, 295–319.

16. ——: 1986, "Help for the Good Samaritan Paradox," *Philosophical Studies* 50, 117–127.
17. ——: 1988, "*Prima facie* Obligation: Its Destruction and Reconstruction," in Ernest LePore and Robert van Gulick (eds.), *John Searle and his Critics* (Cambridge, Mass.: Basil Blackwell, 1991): 359–370.
18. Rescher, N. and Brandom, R.: 1979, *The Logic of Inconsistence* (Totowa, N.J.: Rowman and Littlefield).
19. Sayre-McCord, G.: 1987, "Conditional Obligations and Morally Irrelevant Considerations: Comments on Belzer and Loewer," American Philosophical Association, Pacific Division, San Francisco, March 26, 1987.
20. Scheffler, S.: 1982, *The Rejection of Consequentialism*, Oxford University Press, Oxford.

MARVIN BELZER
Department of Philosophy
Bowling Green State University
Bowling Green, OH 43403-0222
USA

BARRY LOEWER
Department of Philosophy
Rutgers University
New Brunswick, NJ 08903
USA