A LOGIC OF DELIBERATION

Marvin Belzer

Advanced Computational Methods Center
University of Georgia
Athens, GA 30602

ABSTRACT

Deliberation typically involves the formation of a plan or intention from a set of values and beliefs. I suggest that deliberation, or "practical reasoning," is a form of normative reasoning and that the understanding and construction of reasoning systems that can deliberate and act intentionally presupposes a theory of normative reasoning. The language and semantics of a deontic logic is used to develop a theory of defeasible reasoning in normative systems and belief systems. This theory may be applied in action theory and to artificial intelligence by identifying expressions of values, beliefs, and intentions with various types of modal sentences from the language.

While there have been some investigations of the structure of normative reasoning in deontic logic, Bayesian decision theory, and philosophical action theory and ethics, there does not yet exist a general theory of normative reasoning. Such a theory is necessary for the understanding and construction of decision-making systems that use normative principles and policies to form plans, strategies, and intentions. A general logic of the all-purpose "normative reasoner," or "deliberator," is needed.

Practical reasoning, or deliberation, in which intentions to act are formed from a set of desires and beliefs, may be a form of normative reasoning. Expressions of desires and intentions may be treated as rules (norms) or evaluative judgments (Davidson 1977). There also may be a normative component in belief-systems. It has been suggested, for example, that the rules of thumb that enable a system to form tentative conclusions from incomplete information are expressions of "ratiocinative desires" (Doyle 1983a); and that "epistemic policies" guide an epistemic agent in revising beliefs in the light of new information (Stalnaker 1984). The expressions of desires and policies may be interpreted as norms, and therefore an understanding of normative reasoning would be useful in a theory of reasoning with incomplete or new information.

§1. The structure of normative reasoning.

Several features of normative systems must be respected by any adequate formal representation of normative reasoning. First, some rules are *defeasible*, that is, they are generally valid but may have exceptions. Secondly, there is a fundamental distinction between *prima facie rules* and *all-things-considered normative commitments*. The prima facie rules of a system, together with a set of *facts* or *opinions* determine the system's all-things-considered commitments. Thirdly, for some set of sentences the all-things-considered (a.t.c.) closure should be *non-monotonic*, that is, set s is included in set s* but the a.t.c. closure of s is not included in the a.t.c. closure of s*.

These features of rules may be illustrated simply as follows. Suppose that Nixon told you a secret after you promised to comply with these requirements:

(a) You should not tell the secret to Reagan.

(b) You should not tell the secret to Gorbachev.

(c) You should tell Reagan if you tell Gorbachev.

(d) You should tell Gorbachev if you tell Reagan.

Suppose you break promise (b) by a certain time,

(e) You told the secret to Gorbachev,

and you are trying to decide whether you should tell Reagan. If no other rules or facts are relevant then, to comply with the requests as given, clearly you *should* tell the secret to Reagan, because of rule (c)--and in spite of (a). The prima facie rule (a) is *defeasible* because of (c). After you have told Gorbachev you have an all-things-considered commitment expressed by the rule

(f) You should tell the secret to Reagan.

Rules (a) and (f) conflict, yet correct resolution is possible if we recognize that stipulation (a) is a valid prima facie rule whereas (f) expresses a valid all-things-considered commitment *after* it is settled that you have violated rule (b) by telling Gorbachev.

A prima facie rule may be "defeated," in which case it cannot reliably be used to draw normative conclusions. In the example, after you told the secret to Gorbachev the rule (a) was defeated. To use a prima facie rule in particular circumstances to detach an all-things-considered normative conclusion one needs to know that the prima facie rule is not defeated in those circumstances. If it is not defeated, then it can be used--as in the detachment of (f) from (c) and (e).

It does not appear possible to deal separately with the issues of defeasibility and normative reasoning, for even our simple story cannot be represented satisfactorily without defeasible rules – we cannot for instance replace (a) and (b) by

(a') You should not tell Reagan if you do not
tell Gorbachev

and

(b') You should not tell Gorbachev if you do
not tell Reagan.

without omitting from the analysis the significant fact that telling neither is preferable to telling both.

§2. The Deontic Logic 3-D.

Deontic logic is a branch of modal logic whose main goals are to provide a formal representation of rules – typically it does so with modal operators for "ought" and "permissible" – and to provide a semantics for such expressions. A satisfactory deontic logic must be able to represent the distinction between defeasible prima facie (p.f.) rules and all-things-considered (a.t.c.) rules. Moreover it should not permit the detachment of all-things-considered conclusions from defeated rules; and it must have principles that state when such detachment is acceptable.

The deontic logic **3-D** (Loewer and Belzer 1983) meets these requirements. The *language* of **3-D** is a propositional language containing the unary connectives T and F to which are added two dyadic deontic operators O(-/-) and !(-/-), a necessity operator L, and a dyadic operator U(-,-). The wffs of **3-D** are characterized as follows: (a) propositional variables are wffs, and (b) if P,Q are wffs then the following (in addition to the usual truth functional wffs) are wffs: O(Q/P), !(Q/P), LP, and U(Q,P). These statements may be read informally as follows:

O(Q/P): it ought prima facie to be that Q,
given P.

!(Q/P): it ought all-things-considered to be
that Q, given P.

LP: it is settled that P.

U(Q,P): P determines the normative status of
Q.

For tautology T, let OQ = O(Q/T) and !Q = !(Q/T).

A **3-D** *model structure* is a 6-tuple $(\mathbf{W,T,H,I,\leq,F})$ where **W** is a set of momentary world stages, **T** is the set of natural numbers (the set of times), **H** is a subset of the set of functions from **T** into **W** (these functions are possible histories), **I** is a set (of "perspectives"), $\leq$ is a function from **T** x **H** x **I** into the set of weak orderings on **H**, and **F** is a function from **T** x **H** x **I** into **H**.

Call v=<t,h,i>, for time t, history h, and perspective i a *temporal perspective*. The weak ordering $\leq$v is a ranking of possible histories according to the extent to which the histories comply with the *values* of perspective i at time t in history h (cf. Lewis 1973, 1974). The most highly ranked histories are those at which no value or rule is violated. As one descends the ranking more and/or more serious violations occur. This allows for the interpretation of prima facie rules. O(Q/P) is to hold relative to the temporal perspective v just in case Q is true at each of the most highly ranked P-histories in the p.f. ranking $\leq$v.

The set **F**v is the set of histories accessible at v. For an *objective* interpretation we stipulate that

$$\mathbf{F}(<t,h,i>) = \mathbf{F}(<t,h,i^*>)$$

for all i* (that is, in the objective interpretation the perspective i is not relevant to accessibility).* Let P be *settled* at v just in case P is true at each of the histories in the set **F**v (cf. Thomason, 1970). LP says that P is settled.

Now we want to use the p.f. ranking $\leq$v and the set **F**v to define a new ranking $\leq$'v with which to interpret expressions of all-things-considered (a.t.c.) commitments !(Q/P). The main idea to be used is that the a.t.c. ranking for v can be defined as the ranking that results when all histories that are inaccessible at v are removed from the p.f. ranking for v. Given an ordering x on H and subset y of H, let the *restriction* of x to y be the ordering z that results by removing from x each element of H not in y.** Let $\leq$'v be the restriction of $\leq$v to **F**v. !(Q/P) is to hold at v just in case Q is true at each most highly ranked history in $\leq$'v at which P is true.

An *interpretation* [ ] on a **3-D** model structure is defined as follows: [ ] assigns to each propositional variable a subset of **T** x **H** x **I** where we stipulate that for non-modal P:

$$<h,t,i> \varepsilon [P] \text{ iff for all } t^* \varepsilon \mathbf{T} \text{ and } i^* \varepsilon \mathbf{I}, <h,t^*,i^*> \varepsilon [P].$$

In other words, only histories--and not perspectives or times--are relevant to the evaluation of non-modal propositional variables. Recursion clauses for the truth functional connectives are as expected. Now let [Q/P] be the class of weak orderings $\leq$ on H that are such that:

$$Ej(j \ \varepsilon \ [P\&Q] \text{ and } (k)(k \ \varepsilon \ [P\&{\sim}Q] \rightarrow \text{not}(k \leq j)).$$

In other words, [Q/P] is the class of weak orderings on H in which some P&Q-history is ranked more highly than any P&~Q-history.*** For v=<h,t,i>, and j,k ε H:

$$v \ \varepsilon \ [O(Q/P)] \text{ iff } \leq v \ \varepsilon \ [Q/P].$$

$$v \ \varepsilon \ [!(Q/P)] \text{ iff } \leq' v \ \varepsilon \ [Q/P].$$

$$v \ \varepsilon \ [LP] \text{ iff } \mathbf{F}v \subseteq [P].$$

For U(Q,P) let us say first that for x$\subseteq$H, *f* (v,x) is the set of most highly ranked histories in x according to $\leq$v. Also for x,y $\subseteq$H, let

* Cf. §5 below for a *subjective* interpretation of L that does depend on i.

** For example, suppose that H = {1,2,3,4,5} and x is the ordering

(4,5) < (1,2,3)

and y = {1,2,4}. The *restriction* z of x to y would be the ordering

4 < (1,2).

*** The *proposition* expressed by P sometimes is identified with the set of histories [P]. Analogously, the class of rankings [Q/P] may be identified with the *norm* expressed by the sentence "it ought to be that Q, given P," which contains no p.f. or a.t.c. qualifiers.

$$x =P y$$

say

$$x \subseteq [P] \text{ iff } y \subseteq [P] \text{ and } x \cap [P] \neq \Lambda \text{ iff } y \cap [P] \neq \Lambda.$$

The recursion clause for U(Q,P) is as follows:

$$v \in [U(Q,P)] \text{ iff } (x)(x \subseteq \mathbf{H} \ \& \ \mathbf{F}v \subseteq x. \rightarrow f(v,[P] \cap x) =Q\ f(v,[P])).$$

This clause guarantees that if

$$v \in [O(Q/P)]$$

and

$$v \in [U(Q,P)]$$

then there is no R such that v ε [LR] and

$$v \in [\sim O(Q/P\&R)].$$

In reasoning on the basis of defeasible principles of the form O(Q/P), U(Q,P) plays the role of asserting roughly that "other things are equal" – more precisely, that relative to what is settled, P determines the normative status of Q.

The logic of both O(-/-) and !(-/-) is **CD** (van Fraasen 1972, Lewis 1974). The logic of the objective L is **S4**. The question of a complete proof theory for **3-D** remains open. Here are some important formulas that are valid in **3-D**:

(1) $O(Q/P) \ \& \ LP \ \& \ U(Q,P) \rightarrow\ !Q.$

(2) $O(Q/P\&R) \ \& \ U(Q,P\&R) \ \& \ LP \rightarrow\ !(Q/R).$

(3) $\sim O(Q/P\&R) \ \& \ U(\sim Q,P\&R) \ \& \ LP \rightarrow \sim!(Q/R).$

(4) $\sim O(Q/P) \ \& \ U(\sim Q,P) \ \& \ \sim!P \rightarrow \sim!Q.$

(5) $O(Q/P) \ \& \ U(Q,P) \rightarrow \sim L\sim Q.$

(6) $O(Q/P) \ \& \ U(Q,P) \rightarrow \sim L\sim P.$

(7) $LQ \rightarrow\ !Q.$

(8) $!Q \rightarrow \sim L\sim Q.$

(9) $U(Q,P) \ \& \ LR \rightarrow (O(Q/P) \equiv O(Q/P\&R)).$

(10) $U(Q,P) \ \& \ U(R,P) \rightarrow U(Q\&R,P).$

(11) $U(Q,P) \ \& \ LR \rightarrow U(Q,P\&R).$

(12) $!(Q/P\&R) \ \& \ LP \rightarrow\ !(Q/R).$

(13) $\sim!(Q/P\&R) \ \& \ LP \rightarrow \sim!(Q/R).$

(14) $!P \ \& \ L(P \rightarrow Q) \rightarrow\ !Q.$

An application of **3-D** can be illustrated with the "promising" example introduced above (for other applications, cf. Loewer and Belzer 1983, 1986; Belzer 1986a). The relevant rules may be represented as

(a#) O~r

(b#) O~g

(c#) O(r/g)

(d#) O(g/r)

where r stands for 'You tell the secret to Reagan' and g for 'You tell the secret to Gorbachev'. Suppose that from your perspective each of these rules is true and that g is settled. If there is no settled c such that ~O(r/g & c), then you are committed to !r. On the other hand, if ~Lg holds then so also does !~r .

A.t.c. closure is non-monotonic in **3-D** in the sense that a.t.c. commitments relative to a set s of p.f. rules and settled propositions may not hold relative to a superset of s. For instance, let s be the set that contains (a#)-(d#) and let s* include s and also contain Lg. !~r is contained in the a.t.c. closure of s but not in the a.t.c of s* even though s* includes s. The non-monotonicity of a.t.c. closure is owing to the *defeasibility* of the p.f. rules, where

> O(Q/P) is *defeasible* at v iff there is
> some R such that ~O(Q/P & R) holds at v.

(cf. Belzer 1985a). A rule O(Q/P) is *defeated* at v iff there is some R such that both ~O(Q/P & R) and LR hold at v. For instance if p is settled at v, then O~g is defeated at v.

To see the role of the U-statements in **3-D**, suppose that O(Q/P) and LP hold at v. We cannot conclude that !Q for for ~O(Q/P&R) and LR may hold at v; if so, O(Q/P) is defeated at v. However we can infer !Q at v if we know both that O(Q/P) and LP hold at v and that U(Q,P) also holds at v, for U(Q,P) guarantees that no proposition that holds at v defeats O(Q/P).

To complicate the example a bit more, suppose also that O(~r/t) is true, for t 'You tell the secret to Thatcher', and that t as well as g is settled. Is !r true now? It depends. O(r/g) and O(~r/t) may be equally important in the relevant system, or one may have more *weight* than the other. Such relationships can be expressed in 3-D, as is shown in the following section.

§3. Conflicts and Relative Weight.

The distinction between prima facie and all-things-considered reasons for an action is familiar to legal, ethical, and action theorists. Philosophers have stressed the importance of this distinction for formal deontic systems. Much of the work in deontic logic is of marginal interest to those concerned with practical reasoning because it ignores problems due to conflicts of prima facie reasons (Raz 1978). **3-D** however is an exception to this claim, for it can be used to represent conflicting prima facie reasons. In a prima facie conflict, both O(Q/P) and O(S/R) are true and

$$(P \ \& \ R \ \& \sim(S \equiv Q))$$

is settled. The metaphorical notion of relative *weight* that is important in the resolution of conflicts can be defined, as follows:

> O(Q/P) has *greater relative weight*
> than O(S/R) iff O(Q/P & R & ~(S ≡ Q))

(cf. Belzer 1985a,b). In the example discussed above, suppose that O(r/g & t) is true; if so, then O(r/g) has greater relative weight than O(~r/t), so !r would hold if O(r/g & t) itself is not defeated. On the other hand, if neither O(r/g & t) nor O(~r/g & t) is true then neither O(r/g) nor O(~r/t) has greater weight than the other, and it is reasonable to suggest that neither !r nor !~r should hold.

The importance of being able to formulate precise expressions of relative weight between rules is that it may make possible a theory of practical reasoning and rational decision-making that does not depend on quantitative utility functions. While implementation of practical reasoning eventually may involve a combination of qualitative rules and numerical evaluation functions, Bayesian statistical decision making has played only a limited role in artificial intelligence. It often is pointed out that it is not easy to apply Bayesian techniques directly because of both the amount of information that must be supplied in the form of conditional probabilities, prior probabilities, and utilities and the awkwardness of modifying the formulation (Doyle 1983b). However, as Ginsberg (1985) suggests, these problems cannot be regarded as conclusive without having compared Bayesian implementations with others based on qualitative rules having differing relative weights.

§4. Applying 3-D in Belief Systems.

Belief systems also may be structured by a distinction that parallels the distinction between p.f. and a.t.c. norms in normative systems, for we can distinguish between p.f. and a.t.c. *expectations*. Expectations may be treated as rules or norms (for example, as expressions of "ratiocinative desires" or "epistemic policies" for belief revision). However, even if one does not accept the idea that there is a normative component in belief systems a **3-D** type semantics **3-Db** can be used as follows to interpret defeasible reasoning in belief systems.

Let the language of **3-Db** be a propositional language containing T and F, and two dyadic belief operators B(-/-) and L(-/-), a monadic belief operator S, and a dyadic operator V(-/-).

B(Q/P): Q is p.f. expected, given P.

L(Q/P): Q is a.t.c expected, given P.

SP: it is certain that P.

V(Q,P): P determines the doxastic status of Q.

For tautology T, let BQ = B(Q/T) and LQ = L(Q/T).

A **3-Db** model structure is a 6-tuple (**W**,**T**,**H**,**I**,**\$**,**G**) where **W**, **T**,**H**, and **I** are as above, **\$** is a function from **T** x **H** x **I** into the set of weak orderings on **H**, and **G** is a function from **T** x **H** x **I** into **H**. For a temporal perspective v=<t,h,i>, let **\$'**v be defined as the restriction of **\$**v to **G**v. In an interpretation [ ] on a **3-Db** model structure we have

$$v\ \varepsilon\ [B(Q/P)] \text{ iff } \$v\ \varepsilon\ [Q/P].$$

$$v\ \varepsilon\ [L(Q/P)] \text{ iff } \$'v\ \varepsilon\ [Q/P].$$

$$v\ \varepsilon\ [SP] \text{ iff } \mathbf{G}v \subseteq [P].$$

For V(Q,P) let *g*(v,x) be the set of most highly ranked histories in x according to \$v.

$$v\ \varepsilon\ [V(Q,P)] \text{ iff } (x)(x \subseteq H \ \&\ \mathbf{G}v \subseteq x. \rightarrow \\ g(v,[P] \cap x) =_Q g(v,[P])).$$

The weak ordering \$v is a ranking according to the p.f. expectations at v. All birds fly and every Quaker is a pacifist at each of the most highly ranked histories in \$v, but non-pacifist Quakers and non-flying birds may be found at lower reaches. The prima facie rules of the form B(Q/P) express defeasible rules of thumb that are used in "the common practice of jumping to conclusions when actions demand decisions but solid knowledge fails" (Doyle 1983a, p.1) – they are used to form tentative expectations about the world in the absence of complete information. On the other hand sentences of the form SQ express one's "solid information." The "tentative expectations" to which one is committed by one's p.f. expectations and solid information may be expressed with sentences of the form L(Q/P) . The distinction between the operators S and L corresponds to the distinction between what one feels certain about and what one "expects" (but may not feel certain about) given one's certainties and p.f. expectations. For instance, one may see that a certain flower beneath a certain is yellow--one is certain about that--while merely "expecting a.t.c." that it grew from the seed one planted in the spring and that it would disappear were one to untie the goat.

To consider an example suppose that you p.f. expect Quakers to be pacifists, so you p.f. expect that Nixon is a pacifist if he is a Quaker,

(e#) B(p/q),

and you are certain that Nixon is a Quaker,

(f#) Sq.

If B(p/q) is not defeated, then you are committed to the a.t.c. expectation Nixon is a pacifist, Lp. Of course B(p/q) might be defeated, as happens *if* also you are certain that Nixon is a republican,

(g#) Se,

while expecting p.f. that Nixon is not a pacifist if he is a Republican,

(h#) B(~p/e),

and if B(~p/e) has greater relative weight than B(p/q), that is,

(i#) B(~p/e & q).

The logic of both B(-/-) and L(-/-) is **CD** while the logic of S is "weak **S5**."* Two important theorems about B (also L) may be used to test the conjecture that **3-Db** is useful as a logic of defeasible expectations in belief systems:

(15) B(Q/P) & ~B(Q/P&R) → B(~R/Q).

(16) B(Q/P) & ~BQ → B(~P/~Q).

(15') L(Q/P) & ~L(Q/P&R) → L(~R/Q).

(16') L(Q/P) & ~LQ → L(~P/~Q).

* **S5** minus the reflexivity axiom, cf. Moore 1983. Even though S is used to express what one feels certain about,
SQ & ~Q
nonetheless is consistent (i.e., S is a doxastic, not an epistemic, operator).

Other valid formulas of **3-Db** include:

(17) $B(Q/P) \,\&\, SP \,\&\, V(Q,P) \rightarrow LQ$.

(18) $B(Q/P\&R) \,\&\, V(Q,P\&R) \,\&\, SP \rightarrow L(Q/R)$.

(19) $\sim B(Q/P\&R) \,\&\, V(\sim Q,P\&R) \,\&\, SP \rightarrow \sim L(Q/R)$.

(20) $\sim B(Q/P) \,\&\, V(\sim Q,P) \,\&\, \sim LP \rightarrow \sim LQ$.

(21) $B(Q/P) \,\&\, V(Q,P) \rightarrow \sim S\sim Q$.

(22) $B(Q/P) \,\&\, V(Q,P) \rightarrow \sim S\sim P$.

(23) $SP \rightarrow LP$.

(24) $LP \rightarrow \sim S\sim P$.

(25) $V(Q,P) \,\&\, SR \rightarrow (B(Q/P) \equiv B(Q/P\&R))$.

(26) $V(Q,P) \,\&\, V(R,P) \rightarrow V(Q\&R,P)$.

(27) $V(Q,P) \,\&\, SR \rightarrow V(Q,P\&R)$.

(28) $L(Q/P\&R) \,\&\, SP \rightarrow L(Q/R)$.

(29) $\sim L(Q/P\&R) \,\&\, SP \rightarrow \sim L(Q/R)$.

(30) $LP \,\&\, S(P \rightarrow Q) \rightarrow LQ$.

Each of the operators B(-/-), L(-/-), and S are "implicit" belief operators (cf. Levesque 1984) – both BQ and LQ for instance may hold at v even though Q is not "actively" or "explicitly" expected by the perspective v. The focus of this section is on the distinction between the differing types of expectations, and yet these or similar distinctions also may be necessary in a theory of "explicit" belief (cf. Nute 1986, Belzer 1986b).

§5. Practical Reasoning.

The **3-D** and **3-Db** systems provide the foundation for a general theory of practical reasoning. Let the language of **3-Dpr** be the combined languages of **3-D** and **3-Db**. The *values* of an agent are expressed by sentences of the form O(Q/P) while B(Q/P), SQ, and L(Q/P) express various types of *beliefs. Plans* are expressed by sentences of the form !(Q/P), whereas an *intention* is a special type of plan (one whose expression !(Q/P) is such that the subject in Q is the reflexive "I-myself"). An intention in which the predicate of Q is qualified by "here and now" is a *volition*, or *immediate intention* (Brand, 1984).

In **3-Dpr** a **3-Db** belief sub-system may be embedded into a **3-D** normative reasoning system by imposing a *subjective* interpretation on the operator L of **3-D**. Recall that the interpretation of L in **3-D** is *objective* if L is interpreted independent of perspectives. If so **3-D** sentences of the form !(Q/P) specify the "objective" a.t.c. commitments of a perspective i (that is, at least, the commitments relative to the settled facts at t and the values of i at t). In practical reasoning, however, we want to represent commitments based not on the settled facts at t but rather on the beliefs--in particular, the a.t.c. expectations--*of* i at t. L is the only operator shared by the languages of **3-D** and **3-Db**, and it is the key to embedding a belief system in a normative reasoning system. We give L a *subjective* interpretation by requiring that LQ ("it is settled that Q") holds at v=<t,h,i> in the **3-Dpr** embedding system iff LQ ("it is a.t.c. expected that Q") holds at v in the **3-Db** belief sub-system. The set of *settled* propositions in the practical reasoning system is to be identified at each time with the set of *a.t.c. expectations* to which the belief sub-system is committed.

A **3-Dpr** model structure is an 8-tuple $(\mathbf{W},\mathbf{T},\mathbf{H},\mathbf{I},\leq,\mathbf{F},\$,\mathbf{G})$ where $\mathbf{W}, \mathbf{T},\mathbf{H},\mathbf{I},\leq$, and $\mathbf{F}$ are as in **3-D** while **$** and **G** are as in **3-Db.** Let

$$\cap\$'v$$

denote the set of most highly ranked histories in the a.t.c.-belief ranking **$'v**. The following condition on **3-Dpr** model structures guarantees that L is interpreted coherently (and it guarantees that a proposition is settled at v just if it is a.t.c.-expected at v):

$$(\mathbf{F\$'}) \qquad \mathbf{F}v = \cap \$'v.$$

Interpretations on **3-Dpr** model structures are as in **3-D** and **3-Db**. The logic of O(-/-), !(-/-), B(-/-), and L(-/-) is **CD**. Weak **S5** is the logic of S (as it is also for the monadic O,!,B, and L). Each of (1)–(30) is included among the valid formulas of **3-Dpr**.

In a "well-balanced" agent there are interesting relationships between expressions of various mental states. **3-Dpr** offers a conception of consistency between an agent's "implicit" values, beliefs (of three types), plans, intentions, and volitions; this is a conception of "internal rationality," that is, rationality independent of what one's values and beliefs happen to be. It may be argued for instance that !(Q/P) expresses an acceptable plan for an agent iff !(~Q/P) is not entailed by the agent's values and beliefs.

Given condition (**F$'**) and the subjective interpretation of L, some validities of **3-D** are counter-intuitive in **3-Dpr**, in particular, (7) and (8). (7) says that the a.t.c.-expectations of the well-balanced agent also are plans. But surely this is not necessarily so, since one may expect things about which one is indifferent. Similarly

(31) $SQ \rightarrow\, !Q$

which also holds in **3-Dpr** is unacceptable for the same reason. It is plausible even to hold that if one is certain that Q then one does *not* rationally plan for Q, that is,

(7') $SQ \rightarrow \sim !Q$

(cf. Feldman 1983, Loewer and Belzer 1986). On the other hand, according to (8) one reasonably plans that Q only if one does not a.t.c. expect ~Q; but this also should fail because sometimes one reasonably acts intentionally to bring about the best even while expecting the worst. It is at least more plausible, however, to hold that if one *is certain* that ~Q then one should not be planning that Q, i.e.,

(32) $!Q \rightarrow \sim S\sim Q$,

which also is valid. The truth condition for !(Q/P) can be revised so that (7) and (8) are rejected, and (7') is validated while (32) is maintained:

$$v \in [!(Q/P)] \text{ iff } \leq'v \in [Q/P] \text{ and } \textit{not } \; Gv \subseteq [P \rightarrow Q].$$

Given this revision each of (1), (2), (13), and (14) also fail, but are replaced by

$$(1')\ O(Q/P)\ \&\ LP\ \&\ U(Q,P)\ \&\ {\sim}SQ \rightarrow\ !Q.$$

$$(2')\ O(Q/P\&R)\ \&\ LP\ \&\ U(Q,P\&R)\ \&\ {\sim}S(R \rightarrow Q) \rightarrow\ !(Q/R).$$

$$(13')\ {\sim}!(Q/P\&R)\ \&\ SP \rightarrow {\sim}!(Q/R).$$

$$(14')\ !P\ \&\ L(P \rightarrow Q)\ \&\ {\sim}SQ \rightarrow\ !Q.$$

The "promising" and "pacifist" examples given earlier can be combined to illustrate an application of **3-Dpr**. Suppose holding for your temporal perspective v the values (a#) - (d#), the certainties (f#) and (g#), and the p.f. expectations (e#), (h#), and (i#); and suppose also that for some odd reason you are certain that if Nixon is not a pacifist then you will tell the secret to Reagan,

$$(j\#)\quad S({\sim}p \rightarrow r).$$

Are you committed by these beliefs and values to telling Gorbachev? Assuming no other beliefs or values are relevant to that question you first can conclude L~p because

$$B({\sim}p/q\&e)\ \&\ S(q\&e)\ \&\ V({\sim}p,q\&e) \rightarrow L{\sim}p$$

is an instance of theorem (17), S(q&e) is entailed by Sq and Se, and V(~p,q&e) holds if (as supposed) no other beliefs and values are relevant. Yet L~p and (j#) together entail Lr, by (30). So indeed !g does hold because

$$O(g/r)\ \&\ Lr\ \ \&\ U(g,r)\ \&\ {\sim}Sg \rightarrow\ !g$$

is in instance of (1'), and both U(g,r) and ~Sg hold given the "no other things are relevant" assumption in the example. Your values and expectations commit you to telling Gorbachev the secret. This is an example of practical reasoning in which tentative a.t.c. expectations first are detached from p.f. expectations and certainties, and secondly a.t.c. commitments are detached from values and the a.t.c. expectations.

§6. Summary

A theory of normative reasoning needs to be able to handle the defeasibility of prima facie rules that, together with facts or beliefs, determine all-things-considered commitments. The semantics of **3-D** characterizes these concepts in a formal system, and a similar system **3-Db** characterizes related concepts in the context of belief. Deliberation, or practical reasoning, may be understood as a form of normative reasoning. The combined languages of **3-D** and **3-Db** thus are useful in expressing the mental states that figure in practical reasoning. The system **3-Dpr** combines the semantics of **3-D** and **3-Db** with the condition **(F$')** which embeds a belief system within a more general normative system. **3-Dpr** offers a conception of consistency among implicit values, beliefs of three types, plans, intentions, and volitions.

Acknowledgments.

I am grateful to Barry Loewer, Donald Nute, Andre Vellino, and Michael Lewis for discussions about the role of defeasible reasoning in normative systems, belief systems, and deliberation.

References.

Belzer, M. 1985a. Normative kinematics (I): a solution to a problem about permission. **Law and Philosophy** 4:257-287.

---. 1985b. Normative kinematics (II): the introduction of imperatives. **Law and Philosophy** 4:377-403.

---. 1986a. Reasoning with defeasible principles. **Synthese** 66:1-24.

---. 1986b. **Reasons as norms in a theory of defeasibility and non-monotonic commitment.** Athens, Ga.: Advanced Computational Methods Center research report 01-0008.

Brand, M. 1984. **Intending and Acting.** Cambridge, Mass.: MIT Press.

Davidson, D. 1978. Intending. In Yirmiaku, Y., ed., **Philosophy of History and Action.** Dordrecht: D. Reidel.

Doyle, J. 1983a. **Some theories of reasoned assumptions.** Pittsburgh: Carnegie-Mellon University Computer Science Department technical report no. CMU CS-83-125.

---. 1983a. What AI should want from the supercomputers. **AI Magazine** 4:33-35,31.

Feldman, F. 1983. Obligations -- absolute, conditioned, and conditional. **Philosophia** 12.

Ginsberg, M. 1985. Does probability have a place in non-monotonic reasoning? **Proc. Ninth IJCAI**, 107-110.

Lewis, D. 1973. **Counterfactuals.** Oxford: Basil Blackwell.

---. 1974. Semantic analyses for dyadic deontic logic. In Stenlund, S., ed., **Logical Theory and Semantic Analysis.** Dordrecht: D. Reidel, 1-14.

Levesque, H. J. 1984. A logic of implicit and explicit belief. **Proc. National Conference on Artificial Intelligence**, 198-202.

Loewer, B. and Belzer, M. 1983. Dyadic deontic detachment. **Synthese** 54:295-319.

---. 1986. Help for the Good Samaritan paradox. To appear in **Philosophical Studies** 50.

Moore, R. 1983. Semantical considerations on nonmonotonic logic. **Proc. Eighth IJCAI**, 272-279.

Nute, D. 1985. **A non-monotonic logic based on conditional logic.** Athens, Ga.: Advanced Computational Methods Center research report 01-0007.

---. 1986. Defeasible reasoning. Athens, Ga.: Advanced Computational Methods Center. To appear.

Raz, J. 1978. Introduction. In Raz, J., ed., **Practical Reasoning.** Oxford: Oxford University Press, 1-17.

Stalnaker, R. 1984. **Inquiry.** Cambridge, Mass.: MIT Press.

Thomason, R. 1970. Indeterminist time and truth value gaps. **Theoria** 36:264-281.

van Fraasen, B. 1972. The logic of conditional obligation. **Journal of Philosophical Logic** 1:417-483.