Prepint of a paper appearing in

Synthese

Volume 108 (1996), pp. 269--307.

Agency and Obligation

John F. Horty

Philosophy Department and

Institute for Advanced Computer Studies

University of Maryland

College Park, MD 20742

(Email: horty@umiacs.umd.edu)

Abstract

The purpose of this paper is to explore a new deontic operator for representing what an agent ought to do; the operator is cast against the background of a modal treatment of action developed by Nuel Belnap and Michael Perloff, which itself relies on Arthur Prior's indeterministic tense logic. The analysis developed here of what an agent ought to do is based on a dominance ordering adapted from the decision theoretic study of choice under uncertainty to the present account of action. It is shown that this analysis gives rise to a normal deontic operator, and that the result is superior to an analysis that identifies what an agent ought to do with what it ought to be that the agent does.

Contents

1 Introduction **1**

2 Branching time **2**

3 Agency **5**

4 Oughts in branching time **12**

5 Ought to do: the Meinong/Chisholm analysis **19**

6 Ought to do: a different analysis **21**

7 Hints at a general theory **29**

A Proofs of propositions **35**

1 Introduction

The purpose of this paper is to define and explore a new deontic operator for representing what an agent ought to do, a notion that must be distinguished from that of what ought to be. This new operator is cast against the background of a modal analysis of agency developed by Nuel Belnap and Michael Perloff in a series of papers beginning with [3]. The general approach to agency set out in these papers—which itself relies on a theory of indeterministic time due to Arthur Prior—is sometimes described as *stit semantics*, because it concentrates on a construction of the form 'α (an agent) sees to it that A', usually abbreviated simply as [α *stit*: A]. The goal is to provide a precise semantic account of this *stit* operator within the overall logical framework of indeterminism.

As it happens, Prior's indeterministic temporal framework allows also for the introduction of a standard deontic operator $\bigcirc$, meaning 'It ought to be that ...'. It is natural, therefore, to explore the interactions between this standard deontic operator and the *stit* operator representing agency; and it may seem reasonable to propose a logical complex of the form $\bigcirc$[α *stit*: A]—meaning 'It ought to be that α sees to it that A'—as an analysis of the notion that seeing to it that A is something α ought to do. The motive for this analysis, of course, is a philosophical thesis, advanced by some but disputed by others, according to which what an agent ought to do can be identified with what it ought to be that the agent does; a proposal based on this identification was investigated in [14], and defended there against certain objections found in the literature.

In the present paper, I set out a new and powerful objection to the general idea of identifying what an agent ought to do with what it ought to be that he does; and driven by this objection, I propose a new analysis of what an agent ought to do. This new analysis is based on a loose parallel between action in indeterministic time and choice under uncertainty, as it is studied in decision theory. Very roughly, a particular preference ordering—a kind of dominance ordering—is adapted from the study of choice under uncertainty to the present account of action; it is then proposed that an agent ought to see to it that A whenever the agent has available some action which guarantees the truth of A, and which is not dominated by another action that does not guarantee the truth of A. The primary technical point of the paper is the demonstration that this new analysis of what an agent ought to do gives rise to a normal deontic operator.

The paper is organized as follows. Sections 2 first reviews the theory of indeterministic time. Against this background, Section 3 then develops a particularly simple version of stit semantics, and Section 4 defines a standard deontic operator representing what ought to be. Section 5 combines this standard ought operator with the simple stit operator to yield a representation of what it ought to be that an agent does, and then sets out the hypothesis that this notion can be taken as an analysis also of what an agent ought to do. Section 6 is the heart of the paper: it sets out the objection to this previous analysis, introduces a relation of dominance among actions, and then uses this dominance relation to define a deontic operator that captures a new analysis of what an agent ought to do. Finally, Section 7 describes two ways in which this analysis might be generalized: first, by focusing on strategies of action over time, rather than single actions; and second, by exploring preference criteria other than the simple dominance ordering considered here.

2 Branching time

The theory of indeterminism underlying the present work—introduced in Chapter 7 of Prior's [19], and developed in more detail by Richmond Thomason in [22] and [24]—is based on a picture of moments as ordered into a treelike structure, with forward branching representing the openness or indeterminacy of the future and the absence of backward branching representing the determinacy of the past.

Such a picture leads, formally, to a notion of branching temporal frames as structures of the form $\langle \mathit{Tree}, < \rangle$, in which *Tree* is a nonempty set of moments and $<$ is an ordering on *Tree* that is transitive, irreflexive, and that satisfies the treelike property according to which, for any m_1, m_2, and m_3 in *Tree*, if $m_1 < m_3$ and $m_2 < m_3$, then either $m_1 = m_2$ or $m_1 < m_2$ or $m_2 < m_1$. A maximal set of linearly ordered moments from *Tree* is a *history*, representing some complete temporal evolution of the world. If m is a moment and h is a history, then

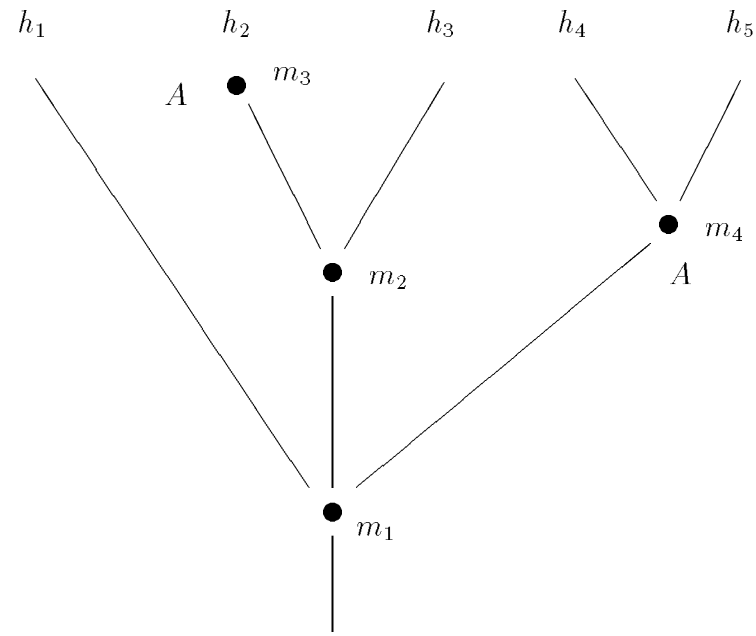

Figure 1: Branching time: moments and histories.

the statement that $m \in h$ can be taken to mean that m occurs at some point in the course of the history h, or that h passes through m. Of course, because of indeterminism, a single moment might be contained in several distinct histories: we let $H_m = \{h : m \in h\}$ represent the set of histories passing through m, those histories in which m occurs.

These ideas can be illustrated as in Figure 1, where the upward direction represents the forward direction of time. This diagram depicts a branching temporal frame containing five histories, h_1 through h_5. The moments m_1 through m_4 are highlighted; and we have, for example, $m_2 \in h_3$ and $H_{m_4} = \{h_4, h_5\}$.

In evaluating formulas against the background of these branching temporal frames, it is a straightforward matter to define a notion of truth at a moment adequate for the truth functional connectives, and even for the operator P representing simple past tense: the definitions from standard (linear) tense logic suffice. Since these frames allow alternative possible futures, however, it is not so easy to understand the operator F, representing future tense. Returning again to Figure 1, suppose that, as depicted, the formula A is true at m_3 and at m_4, but nowhere else. In that case, what truth value should be assigned to $\mathsf{F}A$ at the moment m_1?

On the approach advocated by Prior and Thomason, there is just no way to answer this question. Evidently, $\mathsf{F}A$ is true at m_1—A really does lie in the future—if one of the histories h_2, h_4 or h_5 is realized; but it is false on the histories h_1 and h_3. And since, at m_1, each of these histories is still open as a possibility, that is simply all we can say about the situation. In general, in the context of branching time, a moment alone does not seem to provide enough information for evaluating a statement about the future; and what Prior and Thomason suggest instead is that a future tensed statement must be evaluated with respect to a more complicated index consisting of a moment together with a history through that moment. We let m/h represent such an index: a pair consisting of a moment m and a history h from H_m.

Since future tensed statements are to be evaluated at moments and histories together, semantic uniformity suggests that other formulas must be evaluated at these more complicated indices as well. We therefore define branching temporal models as structures of the form $\mathcal{M} = \langle \mathcal{F}, v \rangle$, in which $\mathcal{F}$ is a branching temporal frame and v is a valuation function mapping each propositional constant from the background language into the set of m/h pairs at which, intuitively, it is thought of as true. Where $\models$ represents, as usual, the relation between an index belonging to some model and the formulas true at that index, the base case of the truth definition for branching temporal models tells us simply that propositional constants are true where v says they are:

- $\mathcal{M}, m/h \models A$ if and only if $m/h \in v(A)$ for A a propositional constant.

And the definition extends to truth functions, past, and future as follows:

- $\mathcal{M}, m/h \models A \wedge B$ if and only if $\mathcal{M}, m/h \models A$ and $\mathcal{M}, m/h \models B$,
- $\mathcal{M}, m/h \models \neg A$ if and only if $\mathcal{M}, m/h \not\models A$,
- $\mathcal{M}, m/h \models \mathsf{P}A$ if and only if there is an $m' \in h$ such that $m' < m$ and $\mathcal{M}, m'/h \models A$,
- $\mathcal{M}, m/h \models \mathsf{F}A$ if and only if there is an $m' \in h$ such that $m < m'$ and $\mathcal{M}, m'/h \models A$.

As usual, we say that a formula is *valid* in some class of models if it is true at each index—in this case, each m/h pair—of every model belonging to that class.

It is easy to see that, as long as we confine ourselves to P, F, and truth functional connectives, the validities generated by this definition in branching temporal models coincide with those of ordinary linear tense logic, for the evaluation rules associated with these operators never look outside the (linear) history of evaluation. However, the framework of branching time allows us to supplement the usual temporal operators with an additional concept of settledness, or historical necessity, along with its dual concept of historical possibility. Here, $\Box A$ is taken to mean that A is settled, or historically necessary; $\Diamond A$, that A is still open as a possibility. The intuitive idea is that $\Box A$ should be true at some moment if A is true at that moment no matter how the future turns out, and that $\Diamond A$ should be true if there is still some way the future might evolve that would lead to the truth of A. The evaluation rule for historical necessity is straightforward:

- $\mathcal{M}, m/h \models \Box A$ if and only if $\mathcal{M}, m/h' \models A$ for all $h' \in H_m$;

and $\Diamond A$ can then be defined in the usual way, as $\neg\Box\neg A$.

It is convenient to incorporate this concept of settledness also into the metalanguage: we will say that A is *settled true* at a moment m in a model $\mathcal{M}$ just in case $\mathcal{M}, m/h \models A$ for each h in H_m, and that A is *settled false* at m just in case $\mathcal{M}, m/h \not\models A$ for each h in H_m.

Once the standard temporal operators are augmented with these concepts of historical necessity and possibility, the framework of branching time poses some technical challenges not associated with standard tense logics, but it is also directly applicable to a number of the philosophical presented by indeterminism. Details concerning both the technical issues surrounding branching time and its philosophical applications can be found in Thomason [24]; a more recent discussion of indeterminism occurs in Belnap and Green [2].

3 Agency

We now turn to the treatment of agency within this framework of branching time. Although we follow Belnap and Perloff [3] in its general approach, the particular account set out here differs in detail, resulting in a stit operator that is simpler than that of Belnap and Perloff and for certain purposes more natural. The present account derives most immediately from [14].

Agents and choices

The idea that an agent α sees to it that A is taken to mean that the truth of the proposition A is guaranteed by an action or choice of α. In order to represent this idea, then, we must be able to speak of individual agents, and also of their actions or choices; and so the basic framework of branching time is supplemented with two additional primitives, both drawn from [3].

The first is simply a set *Agent* of agents, individuals thought of as making choices, or acting, in time.

Now what is it for one of these agents to act, or choose, in this way? We idealize by ignoring any intentional components involved in the concept of action, by ignoring vagueness and probability, and also by treating actions as instantaneous. In this rarefied environment, the idea of acting or choosing can be thought of simply as constraining the course of events to lie within some definite subset of the possible histories still available. When Jones butters the toast, for example, the nature of his action, on this view, is to constrain the history to be realized so that it must lie among those in which the toast is buttered. Of course, such an action still leaves room for a good deal of variation in the future course of events, and so cannot determine a unique history; but it does rule out all those histories in which the toast is not buttered.

Our second additional primitive, then, is a device for representing the constraints that an agent is able to exercise upon the course of history at a given moment, the actions or choices open to him at that moment. Formally, these constraints are encoded through a choice function, mapping each agent α and moment m into a partition $Choice^m_\alpha$ of the histories H_m through m; and the idea is that, by acting at m, the agent α is able to determine a particular one of the equivalence classes from $Choice^m_\alpha$ within which the future course of history must then lie, but that this is the extent of his influence. Of course, in order for this choice information to make any sense, we must require that any two histories in H_m that

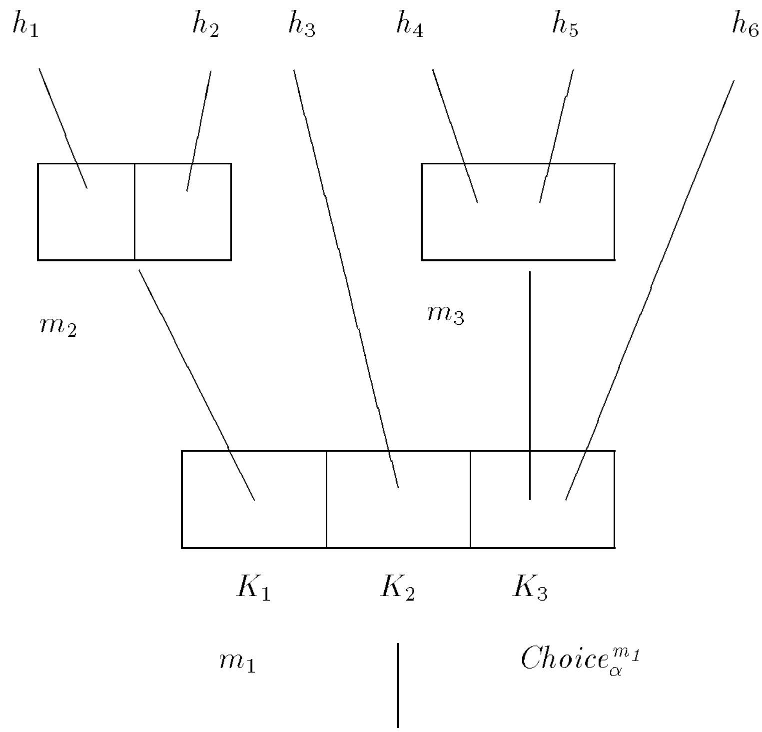

Figure 2: An agent's choices.

have not yet divided at m must lie within the same choice partition; the choices available to an agent at m should not allow a distinction between two histories that do not divide until some later moment.

If K is an equivalence class belonging to $Choice_\alpha^m$, we speak of K as one of the *actions*, or choices, available to α at m; and we let $Choice_\alpha^m(h)$ (defined only when $h \in H_m$) represent the particular action or choice from $Choice_\alpha^m$ that contains the history h. If K is one of the actions available to α at m, we say that that α *performs* the action K at the index m/h just in case h is a history belonging to K. It is important to notice that, as in the evaluation of the future tense, all of the information provided by a full index is required in determining whether an agent performs an action: it makes no sense to say that an agent performs an action at a moment, but only at a moment/history pair. Finally, we speak of the histories belonging to an action K as the possible outcomes that might result from performing this action.

These concepts relating to choice functions can be illustrated as in Figure 2, which depicts a frame containing six histories, and in which the actions available to the agent α at three moments are highlighted. The cells at the highlighted moments represent the actions available to α at those moments. For example, there are three actions available to α at m_1—$Choice_{\alpha}^{m_1} = \{K_1, K_2, K_3\}$, with $K_1 = \{h_1, h_2\}$, $K_2 = \{h_3\}$, and $K_3 = \{h_4, h_5, h_6\}$. Because h_1 and h_2 are still undivided at m_1, these two histories must fall within the same partition there, and likewise for h_4 and h_5. The particular choice partition containing h_5, for example, is K_3, and so we have $Choice_{\alpha}^{m_1}(h_5) = K_3$.

The agent α faces two choices at m_2, but at at m_3 he effectively has no choice: histories divide, but there is nothing α can do to constrain the outcome. (It may be that the outcome can be influenced by some other agent whose choices are not depicted here; or perhaps it is something that just happens, one of nature's choices.) At such a moment, it would be possible to treat the choice function as undefined for α; but it is easier to treat it as defined but vacuous, placing the entire set of histories through the moment in a single equivalence class.

Returning to the moment m_1, we can say that α performs the action K_1 at the index m_1/h_2, for example, that he performs the action K_2 at m_1/h_3, and that he performs the action K_3 at m_1/h_6. Again: since the agent performs different actions along different histories through the moment m_1, it makes no sense to ask what action he performs at that moment. Finally, we can speak of h_4, h_5, and h_6 as the outcomes that might result from performing the action K_3, for example.

When the basic framework of branching time is supplemented with these additional primitives, the result is a *stit frame* of the form

$$\langle Tree, <, Agent, Choice \rangle,$$

with *Tree* and $<$ as before; and we can define a *stit model* as a structure of the form $\mathcal{M} = \langle \mathcal{F}, v \rangle$, in which $\mathcal{F}$ is a stit frame and v a valuation mapping each propositional constant, as before, into a set of m/h pairs. It is these structures that provide the backdrop for the current treatment of agency; the claim is that the structures are not just mathematical curiosities, but describe—up to a legitimate idealization—the world in which agents act.

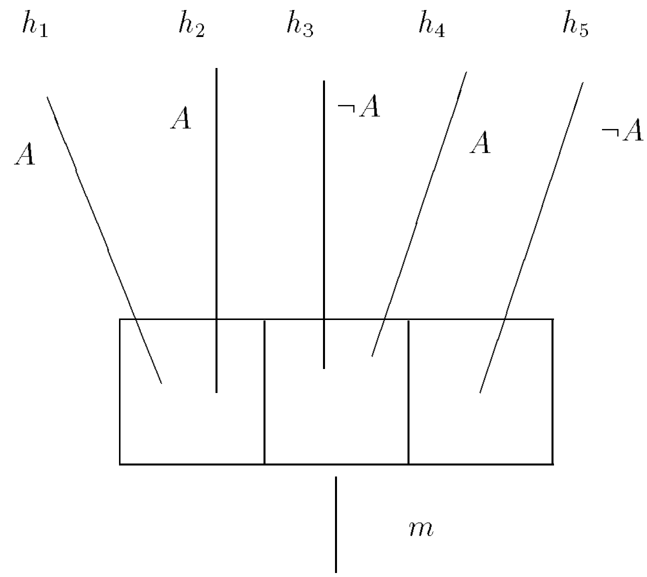

Figure 3: [α *cstit*: A] true at m/h_1.

The Chellas stit

The particular stit operator to be employed in this paper is described in [14] as the "Chellas stit"—and represented there as *cstit*—because it is an analog in the present framework of the agency operator first studied in Brian Chellas's [5].

The idea behind this *cstit* operator is simple: the statement [α *cstit*: A] is to hold at an index m/h just in case α performs an action at m/h that guarantees the truth of A; the action might result in a variety of possible outcomes, but the statement A must be true in each of them. This idea leads to a particularly straightforward evaluation rule:

- $\mathcal{M}, m/h \models [\alpha \; cstit\text{: } A]$ if and only if $\mathcal{M}, m/h' \models A$ for all $h' \in Choice^m_\alpha(h)$.

The rule is illustrated in Figure 3.[1] Here, the statement [α *cstit*: A] is true at m/h_1, because the truth of A is guaranteed by the action that α performs at that index: A holds at each m/h' for each h' belonging to $Choice^m_\alpha(h_1)$. But [α *cstit*: A] is not true at m/h_4, for example. Even though the statement A itself happens to hold at this index, the action that

[1]A convention for interpreting figures: when a formula is written next to some history emanating from a moment, the formula should be taken as true at that moment/history pair. Thus, A should be taken as true at m/h_1 in Figure 3, for example, and $\neg A$ as true at m/h_3.

α performs at m/h_4 does not guarantee the truth of A.

In fact, this *cstit* operator is not the primary focus of [14]. Instead, that paper concentrates on another operator known as the "deliberative stit," represented as *dstit*, and definable through the equivalence

$$[\alpha \; dstit\colon A] \equiv [\alpha \; cstit\colon A] \wedge \neg\Box A.$$

The *dstit* operator has certain advantages over the *cstit* operator in the treatment of agency; for example, as shown in [14], it allows for an attractive analysis of the notion of refraining from an action. Nevertheless, the *cstit* operator is simpler and more transparent, and it will be best to concentrate on this operator in the present paper. To illustrate its simplicity, we note that the *cstit* operator supports the principles

$$\begin{array}{ll} RE. & A \equiv B \quad / \quad [\alpha \; cstit\colon A] \equiv [\alpha \; cstit\colon B], \\ N. & [\alpha \; cstit\colon \top], \\ M. & [\alpha \; cstit\colon A \wedge B] \supset [\alpha \; cstit\colon A] \wedge [\alpha \; cstit\colon B], \\ C. & [\alpha \; cstit\colon A] \wedge [\alpha \; cstit\colon B] \supset [\alpha \; cstit\colon A \wedge B], \\ T. & [\alpha \; stit\colon A] \supset A, \\ 5. & \neg[\alpha \; cstit\colon \neg A] \supset [\alpha \; cstit\colon \neg[\alpha \; cstit\colon \neg A]]; \end{array}$$

and is thus an *S5* modal operator.[2] By contrast, the *dstit* operator does not even satisfy the analogue to M, let alone 5.

Ability

One benefit of employing either the *cstit* or the *dstit* operator in the analysis of agency is that, in either case, a natural treatment of ability lies close at hand. We can assume in either case that an agent's ability (personal can-do) can be represented through a simple combination of ordinary historical possibility (impersonal can) together with the appropriate stit operator (personal to-do). In the present context, the result is an analysis according to which the formula

$$\Diamond[\alpha \; cstit\colon A]$$

[2] The labels for these principles are drawn from Chellas [6].

can be taken to expresses the claim that α is able to see to it that A.

This style of analysis runs contrary to a well-known thesis of Anthony Kenny's, who argues in [15] and [16] that the logic of ability cannot be formalized using the techniques of modal logic. Kenny follows G. H. von Wright in describing the 'can' of ability as a dynamic modality, and puts the point as follows: "ability is not any kind of possibility; ...dynamic modality is not a modality" [16, p. 226].

The central thrust of Kenny's argument is directed against attempts to represent the 'can' of ability as a possibility operator in a modal system with the usual style of possible worlds semantics. Kenny claims that attempts along these lines are doomed to failure: any natural possibility operator, he says, must satisfy the two schemata

$$T\Diamond. \quad A \supset \Diamond A,$$

$$C\Diamond. \quad \Diamond(A \vee B) \supset .\Diamond A \vee \Diamond B;$$

and he argues persuasively that the 'can' of ability does not satisfy either of these. As a counterexample to the first, Kenny considers the case in which a poor darts player throws a dart and actually happens, by chance, to hit the bull's eye; although this shows that it is possible for the darts player to hit the bull's eye, it does not seem to establish his ability to do so. As a counterexample to the second, Kenny imagines a card player who, because he is able simply to draw a card, and all the cards are red or black, is able to draw either a red or a black card; it does not follow that he is able to draw a red card, or that he is able to draw a black card.

Our present analysis of ability escapes from this objection of Kenny's. The notion of historical possibility involved in our analysis, as an *S5* operator, does satisfy both $T\Diamond$ and $C\Diamond$. However, it is not this possibility operator alone that is taken to represent ability, but rather a combination of historical possibility and a stit operator; and the combination fails to satisfy the analogous schemata: both

$$A \supset \Diamond[\alpha\ cstit\colon A],$$

$$\Diamond[\alpha\ cstit\colon A \vee B] \supset .\Diamond[\alpha\ cstit\colon A] \vee \Diamond[\alpha\ cstit\colon B],$$

are invalid. We provide a countermodel only to the first, based on Kenny's darts example, and depicted in Figure 4. Here, m is the moment at which α throws the dart; the cells

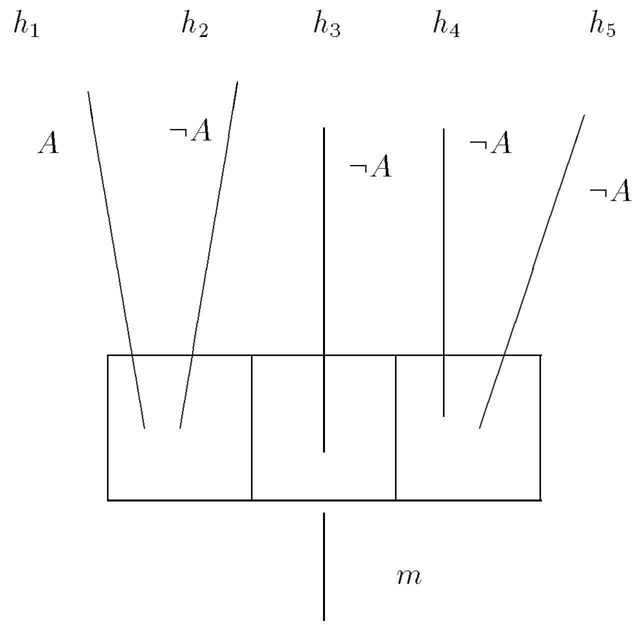

Figure 4: A without $\Diamond[\alpha\ cstit: A]$.

belonging to $Choice_\alpha^m$ represent the possible actions or choices available to α at m; and the formula A means that the dart will hit the bull's eye. Evidently, if the player throws the dart and things evolve along the history h_1, then the dart will hit the bull's eye, but this is not a proposition whose truth the player has the ability to guarantee: although A is true at m/h_1, the formula $\Diamond[\alpha\ cstit: A]$ is not.[3]

4 Oughts in branching time

Standard deontic frames

We begin our treatment of deontic logic by considering a standard way of incorporating the deontic operator $\bigcirc$, representing 'It ought to be that ...', into the framework of branching time.[4] Typically in deontic logic, this ought operator is interpreted against a background

[3]Another response to Kenny's argument from the point of view of modal logic is found in Brown [4]; the relation between the present proposal and Brown's is discussed in [14].

[4]Our presentation follows the approach of Thomason [23]. Work along similar lines, but against the background of a slightly different temporal framework, had previously been carried out by Chellas [5], Montague [18], and Scott [21]; historical details can be found in Thomason [24].

set of possibilities, usually possible worlds. A number of these possibilities are classified as ideal, those in which things turn out as they ought to; and a sentence $\bigcirc A$ is then thought of as true just in case A holds in each of these ideal possibilities—just in case A is a necessary condition for things turning out as they ought to. In the context of branching time, the set of possibilities at a moment m is identified with H_m, the set of histories still available at m; and a nonempty subset of these is taken to represent the ideal histories. A sentence of the form $\bigcirc A$ is then defined as true at an index m/h just in case A is true at m/h' for each history h' from H_m that is classified as ideal.

This picture can be captured formally by supplementing the stit frames described earlier with a function *Ought* mapping each moment m into a nonempty subset $Ought(m)$ of H_m; the result is a *standard deontic* stit frame, a structure of the form

$$\langle Tree, <, Agent, Choice, Ought \rangle,$$

with *Tree*, $<$, *Agent*, and *Choice* as before. Where $\mathcal{M}$ is a standard deontic stit model—a model that results from interpreting our background language against a standard deontic stit frame—the evaluation rule for ought statements can be set out as follows:

- $\mathcal{M}, m/h \models \bigcirc A$ if and only if $\mathcal{M}, m/h' \models A$ for each $h' \in Ought(m)$.

Several logical features of the ought operator developed in this standard way are immediately apparent from the structure of its evaluation rule. First, it is clear that this ought is a normal modal operator—that is, an operator satisfying the principles

$$\begin{array}{ll} RE\bigcirc. & A \equiv B \quad / \quad \bigcirc A \equiv \bigcirc B, \\ N\bigcirc. & \bigcirc\top, \\ M\bigcirc. & \bigcirc(A \wedge B) \supset . \bigcirc A \wedge \bigcirc B, \\ C\bigcirc. & \bigcirc A \wedge \bigcirc B \supset \bigcirc(A \wedge B). \end{array}$$

Second, because the set $Ought(m)$ is nonempty, it is easy to see that the formula $\bigcirc A \supset \Diamond A$ is valid; this formula expresses one version of the characteristic deontic idea that ought imples can—in this case: if it ought to be that A, then it can be that A. Finally, statements of the form $\bigcirc A$, like statements of the form $\Box A$, are always either settled true or settled false.

General deontic frames

Although the study of deontic logics has led the clarification of a number of problems involved in normative reasoning, the topic is often viewed with indifference by researchers interested in ethical theory more generally. Part of the reason for this, I believe, is the impression that these logics are able to model only very crude normative theories—theories that can do no more than classify situations, simply, as either ideal or non-ideal. However, while it is true that standard deontic logics have concentrated on this simple classification of situations, it turns out that the underlying semantic framework can be generalized in a natural way to accommodate a much broader range of normative theories.

In order to arrive at this generalization, in the present context of branching time, let us now imagine that each history through a moment, rather than being classified simply as ideal or non-ideal, is assigned a particular value at that moment. These values, chosen from some general space of values, are to represent the worth or desirability of the histories.

This change in perspective can be effected formally by replacing the primitive *Ought* in the frames described above with a function *Value* that associates each moment m with a mapping of the histories belonging to H_m into the set of values. Depending on the nature of the particular normative theory that is being modeled, the values themselves can be conceived of in different ways, and subjected to different ordering relations; but we will asuume that the space of values is always at least partially ordered by $\leq$, so that $Value_m(h) \leq Value_m(h')$ means that h' has a value at m greater than or equal to that of h. The result can be characterized as a *general deontic* stit frame, a structure of the form

$$\langle Tree, <, Agent, Choice, Value \rangle.$$

In the environment of these new frames, the evaluation rule set out above for ought statements must be abandoned, of course. But it is possible to define a coherent ought operator in this new environment by requiring that a statement of the form $\bigcirc A$ should be true at the index m/h whenever A is true along some history through m, and then true also at every history through m of equal or greater value. Where $\mathcal{M}$ is a general deontic stit model, resulting from the interpretation of our background language against a general deontic stit frame, this idea leads to the following evaluation rule:

- $\mathcal{M}, m/h \models \bigcirc A$ if and only if there is some history $h' \in H_m$ such that (1) $\mathcal{M}, m/h' \models A$, and (2) $\mathcal{M}, m/h'' \models A$ for all histories $h'' \in H_m$ such that $Value_m(h') \leq Value_m(h'')$.

The new rule is similar in spirit to the previous version. In the new environment, we can no longer think of $\bigcirc A$ as true whenever A is a necessary condition for achieving an ideal history, since we are no longer presented with a set of histories classified as ideal; instead, we think of $\bigcirc A$ as true whenever A is a necessary condition for achieving a history whose value is at least as great as some particular value.

It is easy to see that the general deontic framework presented here is, in fact, a conservative generalization of the standard deontic framework set out earlier: any standard deontic stit model can be coded into a general deontic stit model in such a way that the same set of ought statements is supported. Suppose that we allow only the two values 0 and 1, ordered so that $0 \leq 1$. We can then map each standard deontic stit model into a general deontic stit model just like the standard model, except that at each moment it assign the value 1 to those histories that the standard model classifies as ideal, and the value 0 to those histories that the standard model classifies as non-ideal. More exactly, where $\mathcal{M}$ is a standard deontic stit model, we let $\mathcal{M}'$ be a model just like $\mathcal{M}$, except that $Value_m(h) = 1$ in $\mathcal{M}'$ just in case $h \in Ought(m)$ in $\mathcal{M}$, and $Value_m(h) = 0$ otherwise. It is then a simple matter to verify that a statement $\bigcirc A$ will hold in the general deontic model $\mathcal{M}'$ at an index m/h according to our new evaluation rule just in case $\bigcirc A$ holds in the standard model $\mathcal{M}$ at the same index m/h according to the previous evaluation rule.

In addition to encoding the information provided by the standard deontic case, however, general deontic stit frames can be used also to represent normative theories that allow for more than two values, and in which the ordering among values is more complex. The most prominent of these, of course, are utilitarian theories, which take as their space of values a set of utilities usually thought of as isomorphic to the real numbers. In the present context, these theories can be represented through general deontic stit frames in which the function *Value* associates with each history passing through a moment, as its value, a real number

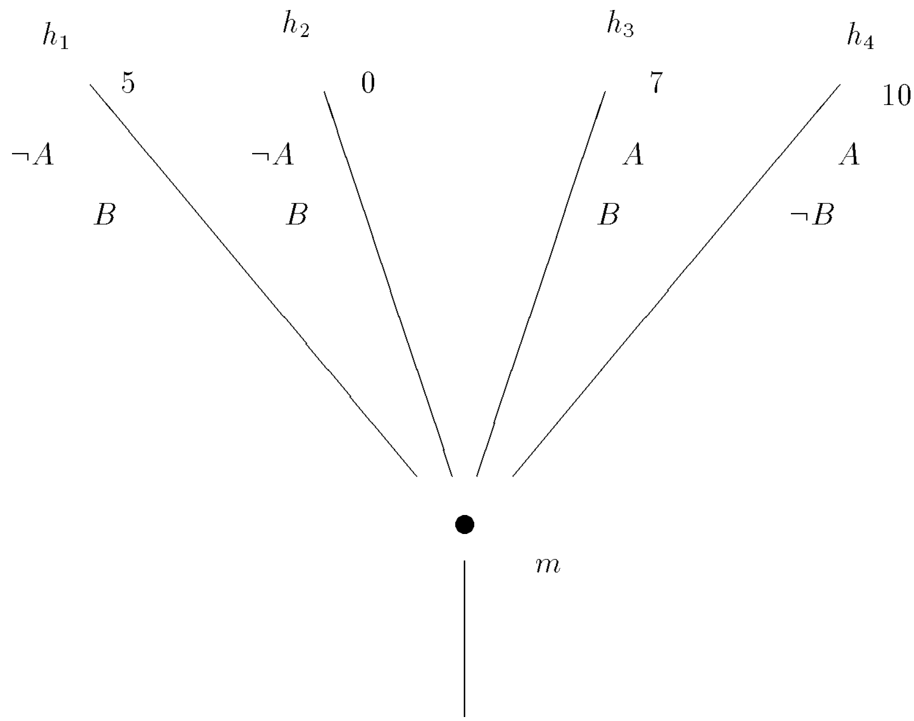

Figure 5: $\bigcirc A$ and $\neg \bigcirc B$.

representing the utility of the history at that moment, and in which the space of values, or real numbers, is subject to its usual ordering. Let us define structures of this kind—general deontic stit frames in which the values are real numbers under their usual ordering—as *utilitarian* stit frames, and the models based on these as utilitarian stit models. We will concentrate on utilitarian models throughout the remainder of the paper.

In depicting these utilitarian stit models, we mark each history through a moment with a number corresponding to its utility at that moment. Thus, Figure 5, for instance, represents a situation in which, at the moment m, the histories h_1, h_2, h_3, and h_4 are taken to possess the utilities 5, 0, 7, and 10, respectively. As a result, we can see that the formula $\bigcirc A$ is settled true at m in this situation, since A holds in h_3 and at each history at least as valuable as h_3. The formula $\bigcirc B$, however, is settled false, since for each history in which B is true, there is a history of equal or greater value in which it is false. A more complicated situation is depicted in Figure 6. Here, we are faced with an infinite number of histories (h_1, h_2, h_3, ...) of ever increasing value (1, 2, 3, ...); the formula A is true at the history h_i when i is

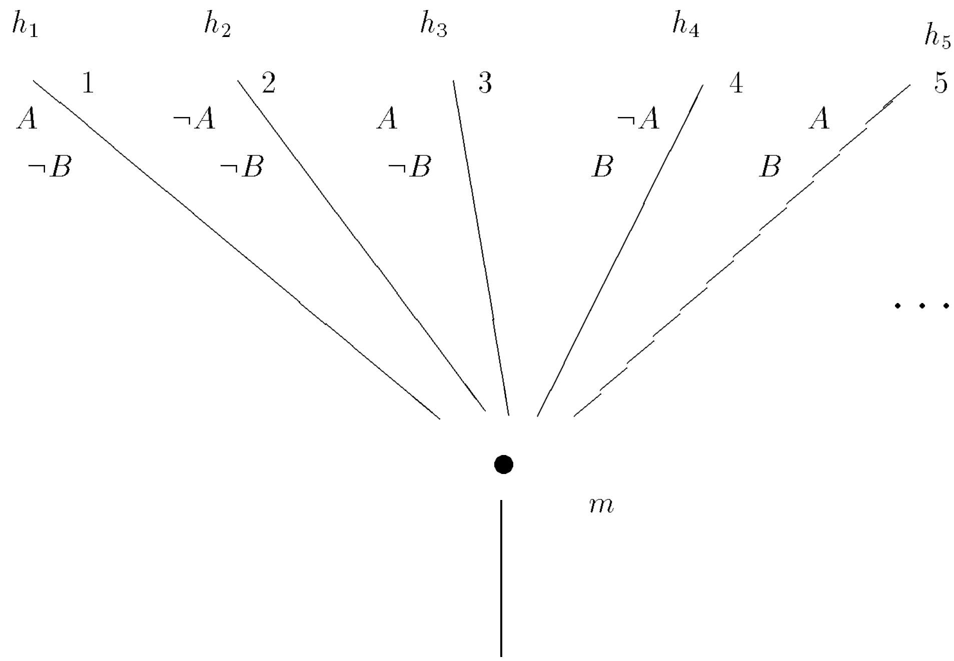

Figure 6: $\neg \bigcirc A$ and $\bigcirc B$.

odd, and false when i is even; and the formula B is true at h_i when i is greater than 3, and otherwise false. As a result, we can see that $\bigcirc A$ is settled false at m, since for each history in which A is true there is a history of equal or greater value in which it is false; but $\bigcirc B$ is settled true, since B is true at all histories at least as valuable as h_4.

Although we concentrate in this paper on utilitarian models, I do wish simply to mention that the general deontic framework developed here is able to accommodate even more radical departures from standard deontic logic. In the utilitarian case, although there are a variety of different values, these values still stand in a linear ordering; but our general framework would allow us to represent theories in which even the assumption of a linear ordering among values is dropped. As an example, consider the approach to deontic logic described by Bas van Fraassen in [25]. Rather than defining oughts against a background set of ideal situations, van Fraassen postulates a background set of imperatives, possibly conflicting; an ought statement is then taken as true if it is is entailed by some maximal consistent subset of these imperatives. Of course, if the background set of imperatives does happen to contain conflicting but individually consistent statements—say, A and $\neg A$—then it will support the truth of conflicting oughts of the form $\bigcirc A$ and $\bigcirc \neg A$.

This idea could be incorporated into the present environment by supposing that each moment m is associated with a separate set $I(m)$ of imperatives—a set of formulas, possibly conflicting, each of which is taken to represent a statement that "ought" to hold at m. Let us now suppose that $Value_m(h)$ is defined as the set of imperatives from $I(m)$ that are true at the index m/h, so that the value assigned to a history at a moment represents the subset of those imperatives operative at that moment that are fulfilled in that history. Since it is better to satisfy more imperatives than fewer, we can take these values as ordered by subset inclusion, so that $Value_m(h) \leq Value_m(h')$ just in case $Value_m(h) \subseteq Value_m(h')$. It then turns out, as noted in [12], that the oughts generated by our new evaluation rule coincide with those supported by van Fraassen's own definition.

Returning to our new deontic evaluation rule, we can see that the ought operator it defines in general deontic stit models shares many of the logical properties of the operator defined by the previous evaluation rule in standard deontic stit models. It should be apparent, for example, that the characteristic deontic formula $\bigcirc A \supset \Diamond A$ is valid in the class of general deontic stit models, and also that any statement of the form $\bigcirc A$ is always either settled true or settled false. In addition, it is easy to verify that the principles $RE\bigcirc$, $N\bigcirc$, and $M\bigcirc$ listed earlier are valid in general deontic models. But if we consider the entire class of general deontic models, then it turns out that the ought operator defined by our new evaluation rule is not a normal modal operator, for the underlying space of values might be ordered in such a way that instances of $C\bigcirc$ are falsified. An example is provided by those general models mentioned above that are designed to represent the theory of van Fraassen [25]: the formula

$$\bigcirc A \wedge \bigcirc \neg A \supset \bigcirc (A \wedge \neg A)$$

will be false at any index at which both $\bigcirc A$ and $\bigcirc \neg A$ are true.

In the present paper, however, we focus on utilitarian stit models, and in these models, the underlying space of values is subject to a linear ordering: for any histories h and h' from H_m, we have either $Value_m(h) \leq Value_m(h')$ or $Value_m(h') \leq Value_m(h)$. It is easy to see that the schema $C\bigcirc$ is valid in any general deontic stit model in which the underlying space of values is subject to a linear ordering; this fact is established as Proposition 1 in the Appendix. Thus, as long as our attention is restricted to the class of utilitarian stit models, the ought operator defined by our new evaluation rule is a normal modal operator, satisfying $C\bigcirc$ as well as $RE\bigcirc$, $N\bigcirc$, and $M\bigcirc$.

5 Ought to do: the Meinong/Chisholm analysis

The utilitarian theory of oughts sketched so far is impersonal, an account of what ought to be. According to this theory, it makes perfect sense to say, for example, that it ought not to snow tomorrow; this means, simply, that there is some history in which it does not snow tomorrow, and that it fails to snow also in any history at least as valuable as that one. There is no implication that anyone ought to see to it that it does not snow, or that anyone can do this. However, just as we analyzed the idea of an agent's personal ability earlier through a combination of ordinary, impersonal possibility and a stit operator, we might hope to arrive at an account of what an agent ought to do in the same way: by combining a stit operator with our impersonal account of what ought to be, we might attempt to analyze what an agent ought to do as what it ought to be that he does.

The idea of analyzing what an agent ought to do as what it ought to be that he does was advanced by a number of Austrian and German writers toward the beginning of the century, notably Meinong and Nicolai Hartmann; and the strategy has been explicitly endorsed by at least one contemporary: Roderick Chisholm suggests in [7, p. 150] that "S ought to bring it about that *p*" can be defined as "It ought to be that S brings it about that *p*."[5] In developing this idea, Chisholm relies on his own treatment of what ought to be, in terms of requirement, and on a simple modal analysis of action that can be found already in the writings of St. Anselm. The same general strategy was studied in some detail in [14], which relied on the *dstit* operator for its treatment of agency, and on the account of what ought to be provided by standard deontic stit models; but in this paper, we will instead employ the *cstit* operator and the more general, utilitarian approach to what ought to be. The result is

[5]Chisholm's paper contains a reference to Hartmann's work; a recent discussion of Meinong's proposal can be found in García [8].

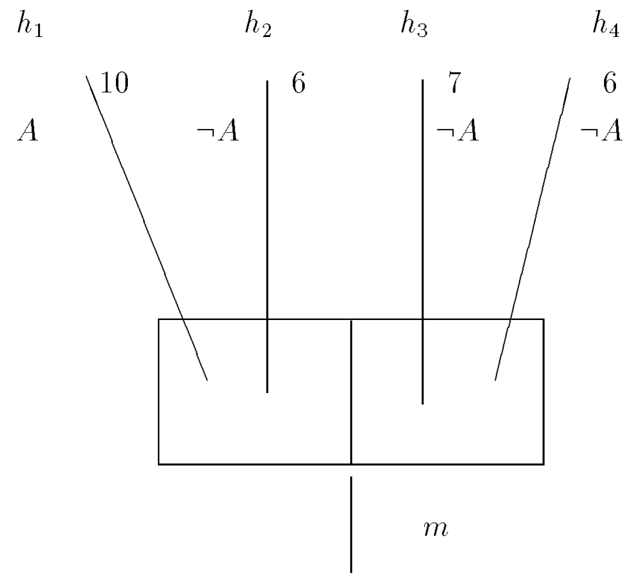

Figure 7: $\bigcirc A$ without $\bigcirc[\alpha\ cstit\colon A]$.

the proposal that a formula of the form

$$\bigcirc[\alpha\ cstit\colon A]$$

can be taken to express the claim that α ought to see to it that A.

This proposal gives us a picture according to which what an agent α ought to do at a particular moment m is determined by the way in which the histories of different value filter through the $Choice^m_\alpha$ partition. Consider, for example, the situation depicted in Figure 7. Here, $\bigcirc A$ is settled true at m. However, $\Diamond[\alpha\ cstit\colon A]$ is settled false: although A ought to hold, there is nothing that α can do about it. Since, as we have seen, any statement of the form $\bigcirc B \supset \Diamond B$ is valid in utilitarian models, we know that

$$\bigcirc[\alpha\ cstit\colon A] \supset \Diamond[\alpha\ cstit\colon A],$$

or that obligation implies ability: whenever it is true that α ought to see to it that A, he must be able to do so. Because α is unable at m to see to it that A, we can thus conclude that $\bigcirc[\alpha\ cstit\colon A]$ is settled false there as well. By contrast, Figure 8 depicts a situation in which $\bigcirc[\alpha\ cstit\colon A]$ is settled true: $[\alpha\ cstit\colon A]$ holds at m/h_1, and also at m/h'' for each history h'' from H_m whose utility is at least as great as that of h_1.

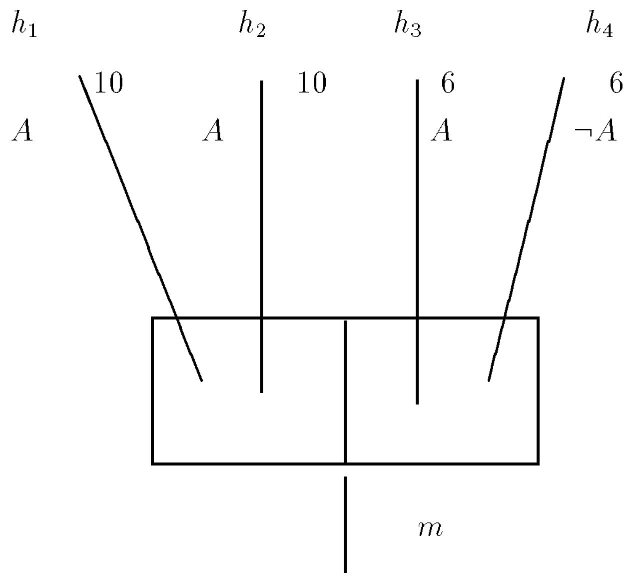

Figure 8: $\bigcirc[\alpha\ cstit: A]$.

Although, according to this analysis, a statement of the form $\bigcirc[\alpha\ cstit: A]$ results from appending the generally applicable operator $\bigcirc$ to the formula $[\alpha\ cstit: A]$, it is still possible to focus on the logically complex connective $\bigcirc[\alpha\ cstit: \ldots]$, and to investigate its properties. It is then easy to see that this complex connective is a normal modal operator, satisfying the principles

$$A \equiv B \quad / \quad \bigcirc[\alpha\ cstit: A] \equiv \bigcirc[\alpha\ cstit: B],$$

$$\bigcirc[\alpha\ cstit: \top],$$

$$\bigcirc[\alpha\ cstit: A \wedge B] \supset . \bigcirc[\alpha\ cstit: A] \wedge \bigcirc[\alpha\ cstit: B],$$

$$\bigcirc[\alpha\ cstit: A] \wedge \bigcirc[\alpha\ cstit: B] \supset \bigcirc[\alpha\ cstit: A \wedge B].$$

6 Ought to do: a different analysis

We have been considering a general approach, advocated by Meinong and Chisholm, according to which what an agent ought to do is determined by what it ought to be that he does. In the present framework, the formula $\bigcirc[\alpha\ cstit: A]$ represents the idea that it ought to be that α sees to it that A, and so according to the Meinong/Chisholm analysis, it is this formula also that represents the idea that α ought to see to it that A.

In fact, this particular version of the Meinong/Chisholm analysis is surprisingly robust:

as shown in [14], it is able to withstand many of the objections advanced by Peter Geach in [9] and Gilbert Harman in [10] and [11, Appendix B] against the general strategy of identifying what an agent ought to do with what it ought to be that he does. Nevertheless, the proposal is vulnerable to another kind of objection. I refer to this objection as the gambler's problem, and use it in this section to motivate a different account of what an agent ought to do.

The gambler's problem

Imagine that an agent α is faced with two options at the moment m: to gamble the sum of five dollars, or to refrain from gambling. If α gambles, we suppose that there is a history in which he wins ten dollars, and another in which he loses and comes away with nothing; but of course, α cannot determine whether he wins or loses. If α does not gamble, we suppose that he preserves his original stake of five dollars no matter how things turn out. Finally, we suppose that the utility associated with each history at m is determined by the sum of money that α possesses in that history. The situation can thus be depicted as in Figure 9. Here, K_1 represents the option of engaging in the gamble, and K_2 the choice of refraining; A represents the statement that α gambles, and h_1 is the history along which α gambles and wins.

It turns out that $\bigcirc[\alpha\ cstit\colon A]$ is settled true at m: the formula $[\alpha\ cstit\colon A]$ is true at m/h_1, and also, trivially, at m/h'' for each history h'' at least as valuable as h_1. The Meinong/Chisholm analysis of what an agent ought to do thus tells us unambiguously that, in this situation, the agent ought to gamble: the most valuable history, with a utility of 10, is that in which he gambles and wins, and it is a necessary condition for achieving this utility that he should gamble. But this is a strange conclusion; for by gambling, the agent runs the real risk of achieving an outcome with the utility of 0, while he is able to guarantee a utility of 5 by refraining from the gamble. From an intuitive point of view, it appears to be impossible to say whether the agent should gamble or not, at least without knowing the odds of winning; and we should be suspicious of any theory that makes a definite recommendation one way or the other.

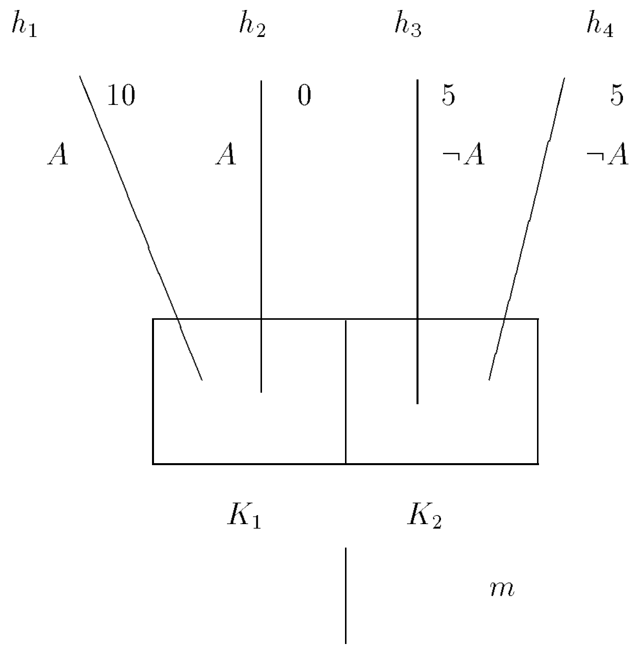

Figure 9: $\bigcirc[\alpha\ cstit: A]$, but not $\odot[\alpha\ cstit: A]$.

The problem presented by this situation seems to reflect a real difficulty with the strategy of identifying what an agent ought to do with what it ought to be that he does. Perhaps it ought to be, in this situation, that the agent gambles; after all, this is what he does in the ideal outcome, the outcome of greatest utility. Still, it does not seem to follow that gambling is something the agent ought to do, since by doing so he risks attaining an outcome of less utility than he could otherwise guarantee.

It might appear that this kind of problem could arise only in a general utilitarian setting, with at least three different values, since it seems to rely upon the possibility that one choice might lead to outcomes both higher and lower in value than the intermediate outcomes resulting from another choice. But a related problem can be seen in a pure deontic setting, which represents outcomes as ideal or non-ideal through the assignment of only two values, 1 and 0. Consider the situation depicted in Figure 10, in which α again has two choices. Again, this situation can be thought of as one in which the agent is faced with accepting or refusing a gamble; and again, K_1 represents the option of engaging in the gamble, and K_2 the choice of refusing the gambling; A represents the statement that α gambles. In this case,

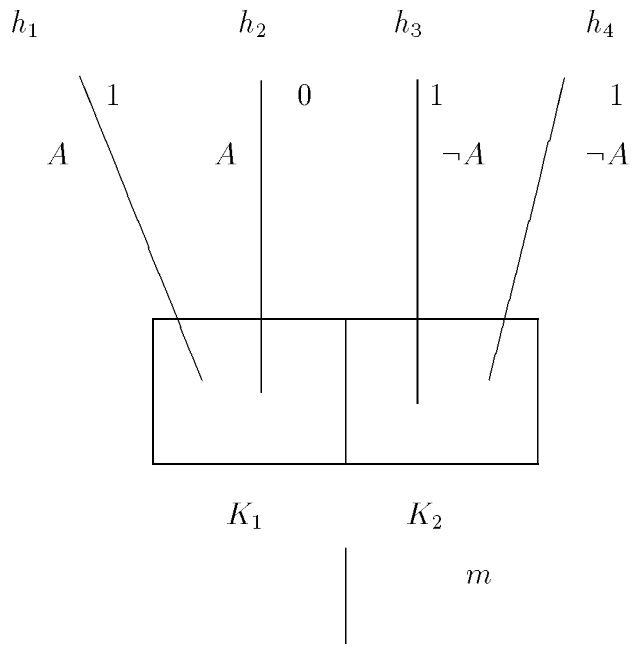

Figure 10: $\neg \bigcirc [\alpha\ cstit\colon \neg A]$, but $\odot[\alpha\ cstit\colon \neg A]$.

however, the gamble is peculiar. If the agent accepts the gamble, we suppose he attains an ideal outcome if he wins, and a non-ideal outcome if he loses; what makes the case peculiar is that, here, the agent can guarantee an ideal outcome by declining the gamble.

It should be obvious that the gamble in this situation is not wise: why should the agent risk a non-ideal outcome simply for the chance of achieving an outcome no greater in value than one that he can guarantee by not gambling at all? Since the gamble is not wise, a correct account of what the agent ought to do should tell us that the agent ought not to gamble in this situation. But this is not the result of the theory that identifies what an agent ought to do with what it ought to be that he does: the statement $\bigcirc[\alpha\ cstit\colon \neg A]$ is settled false at m, since for each history in which the agent refrains from gambling, there is a history of equal value in which he gambles.

Let us return to the situation depicted in Figure 9. One natural way of reacting to situations like this is to ask for additional information—in particular, probabilistic information concerning the various outcomes that might result from the available actions. Suppose that for each action K open to the agent α at m—each K belonging to $Choice^m_\alpha$—we were provided with a probability distribution over the histories belonging to K, where the probability assigned to each history represented its chance of occurring should the agent choose K. We could then define the expected utility of an action K in the usual way, as the sum of the values of the various histories belonging to K, with each value weighted by the the probability assigned to its associated history. This introduction of expected utility would provide us with a linear ordering in value, not only of the histories through a moment, but of the actions themselves that are open to the agent at that moment. And it would then be natural to appeal to this new ordering in defining what an agent ought to do: we could suppose that in any given situation an agent ought to perform some one of those actions open to him whose expected utility is maximal.

The approach just sketched does seem like a promising way to proceed when one possesses the necessary probabilistic information concerning the various outcomes that might result from an agent's actions; but in many situations, this kind of probabilistic information is either unavailable or meaningless. In the literature on decision theory, a situation in which the available actions might lead to their various possible outcomes with known probability is characterized as a case of *risk*; a situation in which the probability with which actions might lead to their various possible outcomes is either unknown or meaningless is characterized as a case of *uncertainty*. Is there anything coherent to say about what an agent ought to do in these cases of uncertainty, when even probabilistic information concerning outcomes is absent?

Dominance

In fact, the decision theoretic treatment of choice under uncertainty describes a variety of ways in which preference orderings can be defined on actions as they are set out there; an introduction to this literature can be found in Chapter 13 of Luce and Raiffa [17]. We proceed in the present paper by adapting a particular one of these preference orderings, a dominance ordering analogous to what Luce and Raiffa call "weak dominance," to the current framework of action. Although the ordering to be defined here ignores the complexities involved in a consideration of independent events and the actions of independent agents—both beyond the scope of the present paper—the simplicity of this ordering makes it especially attractive as a starting point; a refinement designed to accommodate independent events and agents can be found in [13].

Suppose, then, that K and K' are actions open to the agent α at m—members of $Choice_{\alpha}^{m}$—and that they are related as follows: each history belonging to K' is at least as valuable as any history belonging to K, and some history belonging to K' is more valuable than some history belonging to K. In these circumstances, a principle sometimes described as the "sure-thing" principle tells us that K' is a better set of outcomes, a better gamble, than K: by selecting an arbitrary outcome from K', the agent is sure to do at least as well as he would by selecting an arbitrary outcome from K, and he might do better.[6]

Let us now introduce the symbol $\prec$ to represent the preference ordering on actions given by the sure-thing principle. Where K and K' are actions open to an agent at m, we take

$$K \prec K'$$

to mean that: (1) $Value_m(h) \leq Value_m(h')$ for each history h in K and each history h' in K', and (2) $Value_m(h) < Value_m(h')$ for some history h in K and some history h' in K'. When $K \prec K'$, we say that the action K' *dominates* the action K, and we note for future reference that this dominance relation is transitive and asymmetric: if $K \prec K'$ and $K' \prec K''$, we can conclude that $K \prec K''$; and if $K \prec K'$, it is impossible to have $K' \prec K$.

The dominance ordering among actions can be illustrated through our gambling examples. In the situation depicted by Figure 9, we have neither $K_1 \prec K_2$ nor $K_2 \prec K_1$; neither of the actions open to the agent is preferable to the other. In the case of Figure 10, however, we do have $K_1 \prec K_2$, since in that situation it is better for the agent not to gamble.

A new deontic operator

As with expected utilities, this dominance ordering allows us to compare the actions themselves available to an agent at a moment, not merely the histories through that moment.

[6] A discussion of the sure-thing principle from a different perspective can be found in Savage [20, Section 2.7].

This new ordering is weaker than the ordering derived from expected utilities; for example, it is not linear. Nevertheless, the dominance ordering is strong enough to support the definition of a reasonable deontic operator representing what an agent ought to do. Let us introduce the new, two-place operator $\odot[\ldots\ cstit\colon \_\_]$, allowing us to construct statements of the form

$$\odot[\alpha\ cstit\colon A],$$

with the intuitive meaning that α ought to see to it that A. The evaluation rule for this new operator is:

- $\mathcal{M}, m/h \models \odot[\alpha\ cstit\colon A]$ if and only if there is a history $h' \in H_m$ such that (1) $\mathcal{M}, m/h' \models [\alpha\ cstit\colon A]$, and (2) $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h')$ for each history $h'' \in H_m$ such that $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\colon A]$.

And the idea underlying this rule is as follows. The formula $\odot[\alpha\ cstit\colon A]$ is to be true at an index m/h whenever: there is a history h' through m along which $[\alpha\ cstit\colon A]$ is true, hence some action $Choice^m_\alpha(h')$ available to α that guarantees the truth of A, and which is such that, if α does not guarantee the truth of A, it must be that the action he performs is worse than $Choice^m_\alpha(h')$.

Having introduced this new deontic operator directly representing what an agent ought to do, we abandon the Meinong/Chisholm strategy of attempting to explicate what an agent ought to do as what it ought to be that he does. We continue to use the formula $\bigcirc[\alpha\ cstit\colon A]$ as a representation of the idea that it ought to be that α sees to it that A; but the distinct idea that α ought to see to it that A is now carried by the new formula $\odot[\alpha\ cstit\colon A]$.

It should be clear that the new analysis gives us the correct results in our two gambling examples. In the case of Figure 9, where it is appears to be impossible to conclude either that the agent should gamble or that he should not, both $\odot[\alpha\ cstit\colon A]$ and $\odot[\alpha\ cstit\colon \neg A]$ are settled false. In the case of Figure 10, where it seems that the agent should refrain from gambling, the statement $\odot[\alpha\ cstit\colon \neg A]$ is settled true, as desired.

We now turn to some observations concerning the logic of our new deontic operator.

Although perhaps apparent already, it is worth noting explicitly that the notion carried by this new operator of what an agent ought to do is logically neither weaker nor stronger than the notion of what it ought to be that he does, but incomparable: both the formulas

$$\bigcirc[\alpha\ \mathit{cstit}\colon A] \supset \odot[\alpha\ \mathit{cstit}\colon A],$$
$$\odot[\alpha\ \mathit{cstit}\colon A] \supset \bigcirc[\alpha\ \mathit{cstit}\colon A]$$

are invalid in the class of utilitarian stit models. A countermodel to the first is provided by Figure 9; a countermodel to (an instance of) the second is provided by Figure 10. It is interesting to note, however, that if we limit our attention to the class of those utilitarian stit models that can be taken to represent standard deontic stit models—those utilitarian models in which the space of values is limited to 1 and 0, representing the ideal and non-ideal histories—then the first of these two formulas is valid in this more restricted class; the fact is established as Proposition 2 in the Appendix. Thus, while the notion of what it ought to be that an agent does is incomparable in a general utilitarian setting to the notion of what an agent ought ought to do, it is a logically stronger notion in a pure deontic setting.

Even in a general utilitarian setting, however, although the notion of what an agent ought to do is incomparable to the notion of what it ought to be that he does, these two notions are at least guaranteed not to conflict: we will never come across a situation in which the agent ought to see to it that A, although it ought to be that he sees to it that $\neg A$. This guarantee is due to the validity of

$$\neg[\odot[\alpha\ \mathit{cstit}\colon A] \wedge \bigcirc[\alpha\ \mathit{cstit}\colon \neg A]],$$

which is established as Proposition 3 in the Appendix.

It is clear from the structure of the evaluation rule for the new operator that any statement of the form $\odot[\alpha\ \mathit{cstit}\colon A]$ is always either settled true or settled false; and also that the characteristic deontic formula

$$\odot[\alpha\ \mathit{cstit}\colon A] \supset \Diamond[\alpha\ \mathit{cstit}\colon A]$$

is valid. The new operator is, moreover, a normal modal operator, satisfying the principles

$$\begin{array}{ll} RE\odot. & A \equiv B \quad / \quad \odot[\alpha\ cstit\colon A] \equiv \odot[\alpha\ cstit\colon B], \\ N\odot. & \odot[\alpha\ cstit\colon \top], \\ M\odot. & \odot[\alpha\ cstit\colon A \wedge B] \supset . \odot[\alpha\ cstit\colon A] \wedge \odot[\alpha\ cstit\colon B], \\ C\odot. & \odot[\alpha\ cstit\colon A] \wedge \odot[\alpha\ cstit\colon B] \supset \odot[\alpha\ cstit\colon A \wedge B]. \end{array}$$

Although, it is easy to establish $RE\odot$, $N\odot$, and $M\odot$, the verification of $C\odot$ is surprisingly difficult. The reason for this difficulty may become apparent if we recall that the proof of validity provided in Proposition 1 for the analogous formula $C\bigcirc$ relied crucially on the assumption of linearity for the underlying ordering of values. The dominance ordering on actions that figures in the definition of our new deontic operator is not linear; but as Proposition 5 of the Appendix shows, the validity of $C\odot$ can nevertheless be established.

7 Hints at a general theory

The analysis set out here presents, I believe, a coherent theory of what an agent ought to do, and one that improves on the Meinong/Chisholm idea of identifying what an agent ought to do with what it ought to be that he does. There are, however, a number of ways in which the theory as as it stands might be refined and generalized: I close simply by mentioning two of these.

Strategies

First, the theory as it stands focuses only on a moment. It specifies what an agent ought to do at a moment entirely on the basis of the actions available to the agent at that very moment, ignoring any actions that might be available later on. Of course, agents do not usually confine their attention to momentary actions; more often, they work out plans of action over intervals of time. Nevertheless, although a full account of what an agent ought to do over some period of time would ultimately have to be richer than the momentary theory presented here, it might seem that we could safely ignore the problems involved in developing such a full account if we were willing to settle, at first, only for an accurate

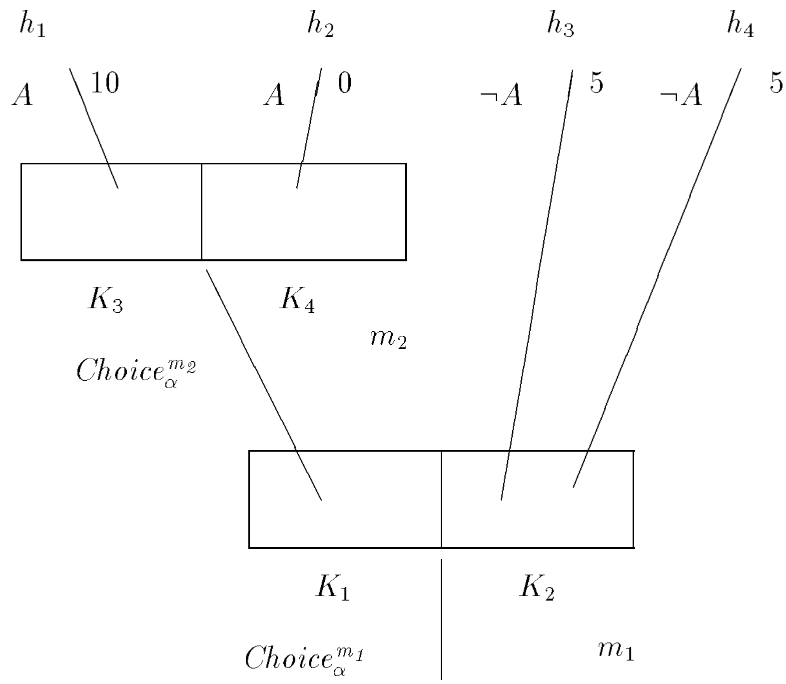

Figure 11: $\odot[\alpha\ cstit: A]$ settled false.

momentary account of what an agent ought to do. Unfortunately, this is not so: there are situations in which, by concentrating only on a moment, ignoring the later actions available to an agent, we are left with a distorted picture of what the agent ought to do even at that very moment.

An example is provided by Figure 11, which depicts the options open to the agent α at the moment m_1, and then also at the later moment m_2. At m_1, the agent faces a choice between $K_1 = \{h_1, h_2\}$ and $K_2 = \{h_3, h_4\}$, and then at m_2 a choice between $K_3 = \{h_1\}$ and $K_4 = \{h_2\}$. The histories h_1 through h_4 possess the values indicated, relative to both m_1 and m_2; and the statement A is true at the indices m_1/h_1, m_1/h_2, m_2/h_1, and m_2/h_2.

Now what should the agent do at the moment m_1? Well, if we look at m_1 alone, the situation appears to be identical to that depicted in Figure 9, our first gambling example. Neither of the actions K_1 or K_2 dominates the other, and so the theory as it stands cannot recommend either action over the other. As a result, both $\odot[\alpha\ cstit: A]$ and $\odot[\alpha\ cstit: \neg A]$ are settled false at m_1.

From an intuitive point of view, however, this result is incorrect. The current situation is not like the earlier gambling example. In the present case, it is as if the agent could first gamble, and then later on choose whether or not he is to win. If the agent selects the action K_1 at m_1, he then faces at m_2 the further choice between K_3 and K_4. By adopting the strategy of first selecting K_1 and then selecting K_3, he can guarantee an outcome of value 10, the highest value that he can guarantee through any available strategy. Since K_1 is the action that α performs at m_1 in the best strategy available, it appears—from this more general perspective, which involves looking at later moments—that K_1 should be classified as a better action than K_2 even at m_1, and therefore, that at m_1 α ought to see to it that A.

Of course, generalizing the notion of what an agent ought to do in this way—evaluating present actions partly on the basis of later possibilities—would involve formulating precisely the notion of a strategy gestured at in the previous paragraph, and then working this notion into the semantics of a new deontic operator. An appropriate notion of strategy can be found in Belnap [1], but the detailed work involved in adapting this notion to the present deontic setting has not yet been carried out.

Alternative preference criteria

The account set out here of what an agent ought to do exploits the analogy between the present theory of action in an indeterministic setting and the decision theoretic treatment of choice under uncertainty. The account is based on one particular preference criterion studied in decision theory, a dominance criterion: it adapts this criterion to define a preference ordering for the present theory of action, and then appeals to the resulting preference ordering in the definition of a new deontic operator.

Although the particular preference criterion relied upon here—the dominance criterion—seems to be especially attractive, it is not the only preference criterion studied within the theory of decision under uncertainty, and others have their merits. A second way of generalizing the present theory of what an agent ought to do, then, is to explore the results of developing an account like that set out here against the background of some of the other preference criteria found in the theory of decision under uncertainty. Simply to illustrate the kind of generalization involved, we now consider how the theory might be developed against the background of the well-known maximin preference criterion.

A decision problem under uncertainty can be formulated as follows. An individual must choose from among a finite number of actions K_1, K_2, ..., K_m. One of a finite number of states of nature s_1, s_2, ..., s_n obtains, but the individual does not know which, and has no information either about relative probabilities. The outcome of his action depends not only on the particular action he performs, but on the state of nature that obtains; thus, an outcome can be defined as a pair $\langle K_i, s_j \rangle$, with A_i an action and s_j a state of nature. With each such outcome there is associated a real number $u[\langle K_i, s_j \rangle]$, representing its utility. Given this information—the actions, the states of nature, and the utilities of outcomes—the goal of a preference criterion is the definition of an intuitively plausible ranking of the available actions.

The maximin criterion is a particularly conservative preference criterion, which ranks each action in accord with the least favorable outcome that might result from that action. Formally, an action K_i is assigned as its security level $sl[K_i]$ the minimum of the numbers $u[\langle K_i, s_1 \rangle]$, $u[\langle K_i, s_2 \rangle]$, ..., $u[\langle K_i, s_n \rangle]$, representing the utilities of the various outcomes that might result from that action. The available actions can then be ranked in accord with their security levels; and according to maximin theory, an agent should choose some action whose security level is maximum.

The picture found in the study of decision under uncertainty is less general than the picture provided by our present framework of utilitarian stit models, for a number of reasons. One important difference is this: each action in decision under uncertainty is associated with a finite number of possible outcomes, determined by the finite number of possible states of affairs; but the set of possible outcomes associated with an action in the present framework—the set of histories contained in that action—may be infinite.

Because an action in the present framework may allow for an infinite set of possible outcomes, its security level cannot be defined simply as the minimum of the values of its outcomes, for there may be no such minimum. Instead, we define the notion by cases. If K is an action available at the moment m, then either there is a lower bound to the set of values of histories contained in K or not. If so, then $sl[K]$—the security level of K—can be defined as the greatest lower bound of these values:

$$sl[K] = glb\{\,Value_m(h) : h \in K\},$$

where *glb* is a function mapping a set of numbers into its greatest lower bound. If not, then we take

$$sl[K] = -\infty,$$

where $-\infty$ is a special value introduced into our utilitarian system of values, and ordered so that it is strictly less than every real number.

The assignment to each action of a security level allows us to rank the actions in accord with the maximin theory; and this new ranking can then be used in the definition of a new deontic operator, representing what an agent ought to do according to the maximin theory. We let the formula $\bigoplus[\alpha\ cstit\colon A]$ represent the idea that α ought according to the maximin theory to see to it that A, with an evaluation rule as follows:

- $\mathcal{M}, m/h \models \bigoplus[\alpha\ cstit\colon A]$ if and only if there is a history $h' \in H_m$ such that (1) $\mathcal{M}, m/h' \models [\alpha\ cstit\colon A]$, and (2) $sl[Choice^m_\alpha(h'')] < sl[Choice^m_\alpha(h')]$ for each history $h'' \in H_m$ such that $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\colon A]$.

The idea, of course, is that $\bigoplus[\alpha\ cstit\colon A]$ should hold at m whenever there is an action K available to α at m that guarantees the truth of A, and which is such that, if α does not guarantee the truth of A, the action he performs has a security level lower than that of K.

In order to illustrate this new operator, let us return again to our first gambling example, depicted in Figure 9. Here, $sl[K_1]$ is 0 while $sl[K_2]$ is 5; the option of declining the gamble has a higher security level than that of gambling. Because of this, it is easy to see that α ought to refrain from gambling according to the maximin theory: the formula $\bigoplus[\alpha\ cstit\colon \neg A]$ is settled true at m

Turning now to the logic of this maximin operator, it is again obvious that the formula $\bigoplus[\alpha\ cstit\colon A]$ is either settled true or settled false at any moment, and that the characteristic deontic formula

$$\bigoplus[\alpha\ cstit\colon A] \supset \Diamond[\alpha\ cstit\colon A]$$

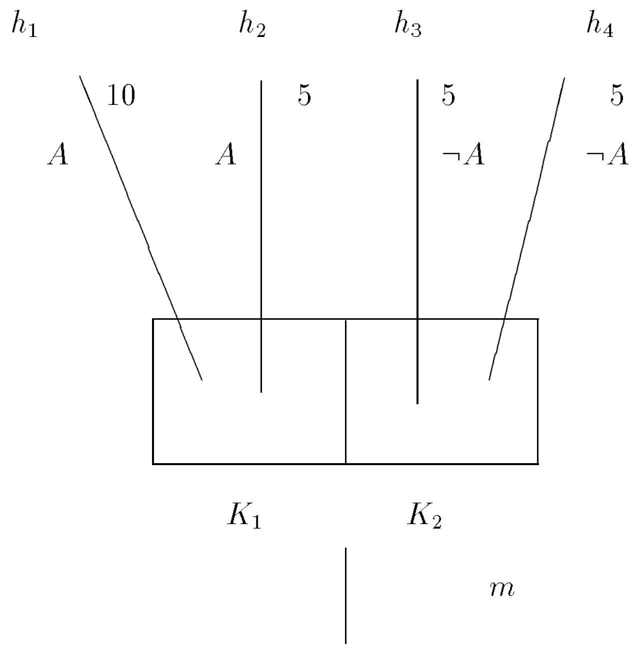

Figure 12: $\odot[\alpha\ cstit\colon A]$, but not $\oplus[\alpha\ cstit\colon A]$.

is valid. Furthermore, we note without proof that this new operator is a normal modal operator, validating $RE\oplus$, $N\oplus$, and $M\oplus$, and $C\oplus$ (the $\oplus$-analogs to $RE\odot$, $N\odot$, and $M\odot$, and $C\odot$). In fact, since the ranking according to security level leads to a linear ordering of actions, the verification of $C\oplus$ is considerably more straightforward than that of $C\odot$.

The maximin conception of what an agent ought to do differs both from our previous analysis of what an agent ought to do, based on the dominance ordering, and also from the notion of what it ought to be that an agent does. Compared to the dominance notion, the maximin conception is logically neither weaker nor stronger: both the schemata

$$\oplus[\alpha\ cstit\colon A] \supset \odot[\alpha\ cstit\colon A],$$
$$\odot[\alpha\ cstit\colon A] \supset \oplus[\alpha\ cstit\colon A]$$

are invalid. As we have seen, a counterexample to (an instance of) the first is provided by Figure 9. A counterexample to the second is found in Figure 12, where $\odot[\alpha\ cstit\colon A]$ is settled true but $\oplus[\alpha\ cstit\colon A]$ is settled false. Still, although these two notions differ, there can be no conflict between them: the validity of

$$\neg[\bigodot[\alpha \; cstit\colon A] \wedge \bigoplus[\alpha \; cstit\colon \neg A]].$$

is easily established.

Compared to the notion of what it ought to be that an agent does, the maximin conception of what the agent ought to do is, again, neither weaker nor stronger; and in fact, here, we do have a real conflict. As we can see from Figure 9, the formula

$$\bigcirc[\alpha \; cstit\colon A] \wedge \bigoplus[\alpha \; cstit\colon \neg A]$$

is satisfiable: although it ought to be that α sees to it that A, what he ought to do according to the maximin conception is to see to it that $\neg A$. This direct conflict between what it ought to be that an agent does and the maximin conception of what the agent ought to do is perhaps not too surprising. For the maximin conception ranks actions entirely on the basis of their worst possible outcomes, completely ignoring any better results to which those actions might lead; the notion of what it ought to be that an agent does, on the other hand, focuses only on the best outcomes that might result from a given action, giving no weight to any risks the agent might have to run in an attempt to achieve those best outcomes.

A Proofs of propositions

Proposition 1 *Let $\mathcal{M}$ be a general deontic stit model in which the underlying space of values is subject to a linear ordering. Then $C\bigcirc$ is true at every index m/h from $\mathcal{M}$.*

Proof Where $\mathcal{M}$ is a general deontic stit model with a linear ordering of values, suppose $\mathcal{M}, m/h \models \bigcirc A \wedge \bigcirc B$. We know from the evaluation rule that there exist histories $h_1, h_2 \in H_m$ such that

> $(*)$ $\mathcal{M}, m/h_1 \models A$, and $\mathcal{M}, m/h'' \models A$ for all histories $h'' \in H_m$ such that $Value_m(h_1) \leq Value_m(h'')$;

and

(∗∗) $\mathcal{M}, m/h_2 \models B$, and $\mathcal{M}, m/h'' \models B$ for all histories $h'' \in H_m$ such that $Value_m(h_2) \leq Value_m(h'')$.

In order to show that $\mathcal{M}, m/h \models \bigcirc(A \wedge B)$, we must show that there is some history $h' \in H_m$ such that (1) $\mathcal{M}, m/h' \models A \wedge B$, and (2) $\mathcal{M}, m/h'' \models A \wedge B$ for all histories $h'' \in H_m$ such that $Value_m(h') \leq Value_m(h'')$. Since the underlying space of values is subject to a linear ordering, we have either $Value_m(h_1) \leq Value_m(h_2)$ or $Value_m(h_2) \leq Value_m(h_1)$. We can thus reason by cases.

Suppose $Value_m(h_1) \leq Value_m(h_2)$. In this case, we identify h' with h_2. Then (∗∗) tells us that $\mathcal{M}, m/h' \models B$, and we can conclude from (∗), since $Value_m(h_1) \leq Value_m(h')$, that $\mathcal{M}, m/h' \models A$. Thus we have (1) $\mathcal{M}, m/h' \models A \wedge B$. Now consider a history $h'' \in H_m$ such that $Value_m(h') \leq Value_m(h'')$. From (∗∗), we know that $\mathcal{M}, m/h'' \models B$. And since $Value_m(h_1) \leq Value_m(h')$, we can conclude that $Value_m(h_1) \leq Value_m(h'')$; and so (∗) tells us also that $\mathcal{M}, m/h'' \models A$. Therefore, $\mathcal{M}, m/h'' \models A \wedge B$; and so we have established that (2) $\mathcal{M}, m/h'' \models A \wedge B$ for all histories $h'' \in H_m$ such that $Value_m(h') \leq Value_m(h'')$.

The argument is symmetric in the case in which $Value_m(h_2) \leq Value_m(h_1)$. ∎

Proposition 2 *Let $\mathcal{M}$ be a utilitarian stit model in which the values assigned to histories are limited to 0 and 1, with* $0 < 1$. *Then the formula* $\bigcirc[\alpha \text{ cstit: } A] \supset \odot[\alpha \text{ cstit: } A]$ *is true at every index m/h from $\mathcal{M}$.*

Proof Let $\mathcal{M}$ be a utilitarian stit model in which the values assigned to histories are limited to 0 and 1, and suppose $\mathcal{M}, m/h \models \bigcirc[\alpha\ cstit\colon A]$. Then there is some history $h_1 \in H_m$ such that (1) $\mathcal{M}, m/h_1 \models [\alpha\ cstit\colon A]$, and (2) $\mathcal{M}, m/h_2 \models [\alpha\ cstit\colon A]$ for all histories $h_2 \in H_m$ such that $Value_m(h_1) \leq Value_m(h_2)$. Now either $Value_m(h') = 0$ for each history $h' \in H_m$ or not. If so, then it follows from (2) that $[\alpha\ cstit\colon A]$ holds at m/h' for each $h' \in H_m$; and so it is easy to see that $\odot[\alpha\ cstit\colon A]$ must be settled true at m. So suppose not—that there is some $h' \in H_m$ such that $Value_m(h') = 1$.

Then by (2) again, we have $[\alpha\ cstit\colon A]$ true at m/h'; and so the first clause is satisfied for the truth of $\mathcal{M}, m/h \models \odot[\alpha\ cstit\colon A]$. Suppose the second clause is not. Then there must be some $h'' \in H_m$ such that (i) $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\text{: } A]$ and (ii) it is not the case that $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h')$. Now either $Value_m(h_3) = 0$ for each history $h_3 \in Choice^m_\alpha(h'')$ or not. If so, then since $h' \in Choice^m_\alpha(h')$ and $Value_m(h') = 1$, it follows from the definition of the $\prec$ relation that $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h')$, contrary to (ii). But if not—if there is some history $h_3 \in Choice^m_\alpha(h'')$ such that $Value_m(h_3) = 1$—then we know from (2) yet again that $\mathcal{M}, m/h_3 \models [\alpha\ cstit\text{: } A]$. But then it is easy to see, since $h_3 \in Choice^m_\alpha(h'')$, that $\mathcal{M}, m/h'' \models [\alpha\ cstit\text{: } A]$ as well, contrary to (i). Hence, the second clause for the truth of $\mathcal{M}, m/h \models \odot[\alpha\ cstit\text{: } A]$ must be satisfied. ∎

Proposition 3 *The formula* $\neg[\odot[\alpha \text{ cstit: } A] \wedge \bigcirc[\alpha \text{ cstit: } \neg A]]$ *is valid in utilitarian models.*

Proof Suppose the contrary, that there is an index m/h in a utilitarian model $\mathcal{M}$ such that $\mathcal{M}, m/h \models \odot[\alpha\ cstit: A]$ and $\mathcal{M}, m/h \models \bigcirc[\alpha\ cstit: \neg A]$. Because $\mathcal{M}, m/h \models \odot[\alpha\ cstit\text{: } A]$, the evaluation rule for this connective tells us that there is a history $h_1 \in H_m$ such that (1) $\mathcal{M}, m/h_1 \models [\alpha\ cstit\text{: } A]$, and (2) $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h_1)$ for each history $h'' \in H_m$ such that $\mathcal{M}, m/h'' \not\models [\alpha\ cstit: A]$. Because $\mathcal{M}, m/h \models \bigcirc[\alpha\ cstit: \neg A]$, we can conclude from the evaluation rule for $\bigcirc$ that there is a history $h_2 \in H_m$ such that (3) $\mathcal{M}, m/h_2 \models [\alpha\ cstit\text{: } \neg A]$, and (4) $Value_m(h'') < Value_m(h_2)$ for each history $h'' \in H_m$ such that $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\text{: } \neg A]$. From (1) we can conclude that $\mathcal{M}, m/h_1 \not\models [\alpha\ cstit\text{: } \neg A]$, and so from (4) that $Value_m(h_1) < Value_m(h_2)$. From (3) we can conclude that $\mathcal{M}, m/h_2 \not\models [\alpha\ cstit: A]$, and so from (2) that $Choice^m_\alpha(h_2) \prec Choice^m_\alpha(h_1)$. But now we have both $Choice^m_\alpha(h_2) \prec Choice^m_\alpha(h_1)$ and $Value_m(h_1) < Value_m(h_2)$; and that is impossible, since the definition of the $\prec$ relation tell us that $Choice^m_\alpha(h_2) \prec Choice^m_\alpha(h_1)$ entails $Value_m(h_2) \leq Value_m(h_1)$. ∎

Proposition 4 *Let K and K' be two actions belonging to $Choice^m_\alpha$, and suppose that neither dominates the other; that is, both $K \prec K'$ and $K' \prec K$ fail. Then either: (A) $Value_m(h) = Value_m(h')$ for all $h \in K$ and $h' \in K'$, or (B) there exist $h \in K$ and $h' \in K'$ such that $Value_m(h) < Value_m(h')$, and there exist $h \in K$ and $h' \in K'$ such that $Value_m(h') < Value_m(h)$.*

Proof Suppose both (A) and (B) are false. The falsity of (A) tells us that $\mathit{Value}_m(h) \neq \mathit{Value}_m(h')$ for some $h \in K$ and $h' \in K'$. The falsity of (B) tells us that either (1) $\mathit{Value}_m(h') \leq \mathit{Value}_m(h)$ for all $h \in K$ and $h' \in K'$ or (2) $\mathit{Value}_m(h) \leq \mathit{Value}_m(h')$ for all $h \in K$ and $h' \in K'$. So suppose, first, that (1). Together with the falsity of (A), however, this yields the result that (3) $\mathit{Value}_m(h') < \mathit{Value}_m(h)$ for some $h \in K$ and $h' \in K'$; but (1) and (3) tell us that $K' \prec K$, contrary to hypothesis. Likewise, if we suppose that (2), the falsity of (A) yields the result that $K \prec K'$, again contrary to hypothesis. Therefore neither (1) not (2) can be assumed along with the falsity of (A); but since either (1) or (2) must hold if (B) is false, the falsity of (B) cannot be assumed along with the falsity of (A). ∎

Proposition 5 *The formula* $C\odot$ *is valid in the class of utilitarian stit models.*

Proof Where $\mathcal{M}$ is a utilitarian stit model, suppose that $\mathcal{M}, m/h \models \odot[\alpha\ \mathit{cstit}\colon A] \wedge \odot[\alpha\ \mathit{cstit}\colon B]$. Then we know that there exist histories $h_1, h_2 \in H_m$ such that

(∗) $\mathcal{M}, m/h_1 \models [\alpha\ \mathit{cstit}\colon A]$ and $\mathit{Choice}^m_\alpha(h'') \prec \mathit{Choice}^m_\alpha(h_1)$ for each history $h'' \in H_m$ such that $\mathcal{M}, m/h'' \not\models [\alpha\ \mathit{cstit}\colon A]$,

and

(∗∗) $\mathcal{M}, m/h_2 \models [\alpha\ \mathit{cstit}\colon B]$ and $\mathit{Choice}^m_\alpha(h'') \prec \mathit{Choice}^m_\alpha(h_2)$ for each history $h'' \in H_m$ such that $\mathcal{M}, m/h'' \not\models [\alpha\ \mathit{cstit}\colon B]$.

In order to show that $\mathcal{M}, m/h \models \odot[\alpha\ \mathit{cstit}\colon A \wedge B]$, we must show that there exists a history $h' \in H_m$ such that (1) $\mathcal{M}, m/h' \models [\alpha\ \mathit{cstit}\colon A \wedge B]$, and (2) and $\mathit{Choice}^m_\alpha(h'') \prec \mathit{Choice}^m_\alpha(h')$ for each history $h'' \in H_m$ such that $\mathcal{M}, m/h'' \not\models [\alpha\ \mathit{cstit}\colon A \wedge B]$. We proceed by cases, with our primary case structure organized around the relation between $\mathit{Choice}^m_\alpha(h_1)$ and $\mathit{Choice}^m_\alpha(h_2)$.

Case I: $\mathit{Choice}^m_\alpha(h_1) \prec \mathit{Choice}^m_\alpha(h_2)$. Here, we identify h' with h_2. We know already from (∗∗) that $\mathcal{M}, m/h_2 \models [\alpha\ \mathit{cstit}\colon B]$. So suppose it were the case that $\mathcal{M}, m/h_2 \not\models [\alpha\ \mathit{cstit}\colon A]$. We could then conclude from (∗) that $\mathit{Choice}^m_\alpha(h_2) \prec \mathit{Choice}^m_\alpha(h_1)$; but since the $\prec$ relation is asymmetric, this would contradict the Case I hypothesis. Thus $\mathcal{M}, m/h_2 \models [\alpha\ \mathit{cstit}\colon A]$.

Combining these observations, we have $\mathcal{M}, m/h_2 \models [\alpha\ cstit\colon A \wedge B]$; and so the first clause is satisfied for the truth of $\mathcal{M}, m/h_2 \models \odot[\alpha\ cstit\colon A \wedge B]$.

In order to see that the second clause is satisfied, we must see that, for each $h'' \in H_m$, if $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\colon A \wedge B]$, then $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h_2)$. So suppose $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\colon A \wedge B]$. We must then have either (i) $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\colon A]$ or (ii) $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\colon B]$. If (ii), then it follows at once from (∗∗) that $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h_2)$. So suppose (i). In that case, it follows from (∗) that $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h_1)$; but then we can conclude that $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h_2)$ from the transitivity of $\prec$ and the Case I hypothesis.

Thus, both clauses are satisfied, and we can conclude in this case that $\mathcal{M}, m/h \models \odot[\alpha\ cstit\colon A \wedge B]$.

Case II: $Choice^m_\alpha(h_2) \prec Choice^m_\alpha(h_1)$. The argument in this case is analogous to that of Case I, with h' chosen as h_1.

Case III: Neither $Choice^m_\alpha(h_1) \prec Choice^m_\alpha(h_2)$ nor $Choice^m_\alpha(h_2) \prec Choice^m_\alpha(h_1)$. We then consider three subcases.

Case III.1: There exists a history $h_3 \in Choice^m_\alpha(h_1)$ such that for all histories $h_4 \in Choice^m_\alpha(h_2)$ we have $Value_m(h_3) \leq Value_m(h_4)$; that is, $Choice^m_\alpha(h_1)$ contains a history whose value is a lower bound of the values of the histories belonging to $Choice^m_\alpha(h_2)$.

Here, we identify h' with h_2. We know from (∗∗) that $\mathcal{M}, m/h_2 \models [\alpha\ cstit\colon B]$. Now suppose it were the case that $\mathcal{M}, m/h_2 \not\models [\alpha\ cstit\colon A]$. We could then conclude from (∗) that $Choice^m_\alpha(h_2) \prec Choice^m_\alpha(h_1)$; but this would contradict the Case III hypothesis. Thus $\mathcal{M}, m/h_2 \models [\alpha\ cstit\colon A]$. Combining these observations, we have $\mathcal{M}, m/h_2 \models [\alpha\ cstit\colon A \wedge B]$; and so, as in Case I, the first clause is satisfied for the truth of $\mathcal{M}, m/h_2 \models \odot[\alpha\ cstit\colon A \wedge B]$.

In order to see that the second clause is satisfied, we must see that, for each $h'' \in H_m$, if $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\colon A \wedge B]$, then $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h_2)$. So suppose again that $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\colon A \wedge B]$. Then as before, we must have must have either (i) $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\colon A]$ or (ii) $\mathcal{M}, m/h'' \not\models [\alpha\ cstit\colon B]$. Again, it follows at once from (ii) and (∗∗) that $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h_2)$. And it follows from (i) and (∗) that $Choice^m_\alpha(h'') \prec Choice^m_\alpha(h_1)$, but here we cannot rely, as in Case I, on transitivity and the case hypothesis to yield $Choice_\alpha^m(h'') \prec Choice_\alpha^m(h_2)$.

Instead, we must note that, according to Proposition 4, the Case III hypothesis can hold under only two conditions: either (A) $Value_m(h) = Value_m(h')$ for each $h \in Choice_\alpha^m(h_1)$ and each $h' \in Choice_\alpha^m(h_2)$; or (B) there exists an $h \in Choice_\alpha^m(h_1)$ and an $h' \in Choice_\alpha^m(h_2)$ such that $Value_m(h) < Value_m(h')$, and there exists an $h \in Choice_\alpha^m(h_1)$ and an $h' \in Choice_\alpha^m(h_2)$ such that $Value_m(h') < Value_m(h)$. Of course, under the condition (A), we can conclude that $Choice_\alpha^m(h'') \prec Choice_\alpha^m(h_2)$ at once from the fact that $Choice_\alpha^m(h'') \prec Choice_\alpha^m(h_1)$.

So suppose condition (B) holds. Then we can conclude as follows that $Choice_\alpha^m(h'') \prec Choice_\alpha^m(h_2)$ from the fact that $Choice_\alpha^m(h'') \prec Choice_\alpha^m(h_1)$ together with the additional information provided by the Case III.1 hypothesis. By this hypothesis, we know that each history from $Choice_\alpha^m(h_2)$ has a value greater than or equal to that of $h_3 \in Choice_\alpha^m(h_1)$; and since $Choice_\alpha^m(h'') \prec Choice_\alpha^m(h_1)$, we know that each history from $Choice_\alpha^m(h'')$ has a value less than or equal that of h_3. Therefore each history from $Choice_\alpha^m(h_2)$ has a value greater than or equal to the value of any history from $Choice_\alpha^m(h'')$. From condition (B), we know that some history from $Choice_\alpha^m(h_2)$ has a value properly greater than that of some history from $Choice_\alpha^m(h_1)$, which must again have a value greater than or equal to that of any history from $Choice_\alpha^m(h'')$. So we know that some history from $Choice_\alpha^m(h_2)$ must have a value properly greater than that of some history from $Choice_\alpha^m(h'')$. Hence we have $Choice_\alpha^m(h'') \prec Choice_\alpha^m(h_2)$.

So both clauses are satisfied, and we can again conclude in this case that $\mathcal{M}, m/h \models \odot[\alpha\ cstit\colon A \wedge B]$.

Case III.2: There exists a history $h_4 \in Choice_\alpha^m(h_2)$ such that for all histories $h_3 \in Choice_\alpha^m(h_1)$ we have $Value_m(h_4) \leq Value_m(h_3)$; that is, $Choice_\alpha^m(h_2)$ contains a history whose value is a lower bound of the values of the histories belonging to $Choice_\alpha^m(h_1)$. The agrument in this case is similar to that of Case III.1, with h' chosen as h_1.

Case III.3: For each history $h_3 \in Choice_\alpha^m(h_1)$ there is a history $h_4 \in Choice_\alpha^m(h_2)$ such that $Value_m(h_4) < Value_m(h_3)$, and for each history $h_4 \in Choice_\alpha^m(h_2)$ there is a history $h_3 \in Choice_\alpha^m(h_1)$ such that $Value_m(h_3) < Value_m(h_1)$; that is, neither $Choice_\alpha^m(h_1)$ not $Choice_\alpha^m(h_2)$ contains a history whose value is lower bound of the values of the histories contained in the other. (To guide imagination, note that this case would be satisfied if each of $Choice_\alpha^m(h_1)$ and $Choice_\alpha^m(h_2)$ contained a history having the value of every real number greater than, say, 4).

Here we can choose h' as either h_1 or h_2; so let us pick h_2. Then just as in Case III.1, we can show that $\mathcal{M}, m/h_2 \models [\alpha\ cstit\colon A \wedge B]$, satisfying the first clause for the truth of $\mathcal{M}, m/h \models \odot[\alpha\ cstit\colon A \wedge B]$. We can continue following the argument of Case III.1 in the treatment of the second clause until it arrives at the intermediate conclusion that $Choice_\alpha^m(h'') \prec Choice_\alpha^m(h_1)$ in case $\mathcal{M}, m/h_2 \not\models \odot[\alpha\ cstit\colon A]$. It is then necessary to conclude from this only that $Choice_\alpha^m(h'') \prec Choice_\alpha^m(h_2)$. But this follows by elementary reasoning from the Case III.3 hypothesis and the definition of the $\prec$ relation.

So both clauses are again satisfied, and we can conclude in this final case that $\mathcal{M}, m/h \models \odot[\alpha\ cstit\colon A \wedge B]$. ∎

Acknowledgments

I wish to thank Nuel Belnap for suggestions and encouragement. This work has been supported by the National Endowment for Humanities through a Fellowship for University Teachers.

References

[1] Nuel Belnap. An austere theory of strategies. Manuscript, Philosophy Department, University of Pittsburgh, 1994.

[2] Nuel Belnap and Mitchell Green. Indeterminism and the thin red line. Manuscript, Philosophy Department, University of Pittsburgh, 1993.

[3] Nuel Belnap and Michael Perloff. Seeing to it that: a canonical form for agentives. *Theoria*, 54:175–199, 1988.

[4] Mark Brown. On the logic of ability. *Journal of Philosophical Logic*, 17:1–26, 1988.

[5] Brian Chellas. *The Logical Form of Imperatives.* PhD thesis, Philosophy Department, Stanford University, 1969.

[6] Brian Chellas. *Modal Logic: An Introduction.* Cambridge University Press, 1980.

[7] Roderick Chisholm. The ethics of requirement. *American Philosophical Quarterly*, 1:147–153, 1964.

[8] J. García. The *tunsollen*, the *seinsollen*, and the *soseinsollen*. *American Philosophical Quarterly*, 23:267–276, 1986.

[9] Peter Geach. Whatever happened to deontic logic? *Philosophia*, 11:1–12, 1982.

[10] Gilbert Harman. Human flourishing, ethics, and liberty. *Philosophy and Public Affairs*, 12:307–322, 1983.

[11] Gilbert Harman. *Change in View: Principles of Reasoning.* The MIT Press, 1986.

[12] John Horty. Deontic logic as founded on nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence*, 9:69–91, 1993.

[13] John Horty. *Agency, Deontic Logic, and Utilitarianism.* Manuscript, University of Maryland, 1994.

[14] John Horty and Nuel Belnap. The deliberative stit: a study of action, omission, ability, and obligation. *Journal of Philosophical Logic*, forthcoming.

[15] Anthony Kenny. *Will, Freedom, and Power.* Basil Blackwell, 1975.

[16] Anthony Kenny. Human abilities and dynamic modalities. In Juha Manninen and Raimo Tuomela, editors, *Essays on Explanation and Understanding: Studies in the Foundations of Humanities and Social Sciences*, pages 209–232. D. Reidel Publishing Company, 1976.

[17] R. Duncan Luce and Howard Raiffa. *Games and Decisions.* John Wiley and Sons, Inc., 1957.

[18] Richard Montague. Pragmatics. In R. Klibansky, editor, *Contemporary Philosophy: A Survey*. Florence, 1968.

[19] Arthur Prior. *Past, Present, and Future.* Oxford University Press, 1967.

[20] Leonard J. Savage. *The Foundations of Statistics.* John Wiley and Sons, 1954. Second revised edition published by Dover Publications, 1972.

[21] Dana Scott. A logic of commands. Manuscript, Philosophy Department, Stanford University, 1967.

[22] Richmond Thomason. Indeterminist time and truth-value gaps. *Theoria*, 36:264–281, 1970.

[23] Richmond Thomason. Deontic logic as founded on tense logic. In Risto Hilpinen, editor, *New Studies in Deontic Logic*, pages 165–176. D. Reidel Publishing Company, 1981.

[24] Richmond Thomason. Combinations of tense and modality. In Dov Gabbay and Franz Guethner, editors, *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, pages 135–165. D. Reidel Publishing Company, 1984.

[25] Bas van Fraassen. Values and the heart's command. *The Journal of Philosophy*, 70:5–19, 1973.