Prepint of a paper appearing in

Journal of Philosophical Logic

Volume 24 (1995), pp. 583--644.

The Deliberative Stit: A Study of Action, Omission, Ability, 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)

Nuel Belnap

Department of Philosophy

University of Pittsburgh

Pittsburgh, PA 15260

(Email: belnap@vms.cis.pitt.edu)

November 23, 1993

Contents

1 Introduction

The idea of treating agency as a modality—representing through an intensional operator the agency, or action, of some individual in bringing about a particular state of affairs—is an old idea, whose roots go back at least as far as St. Anselm, and which has been explored in the present century by a number of writers, including Alan Anderson, Lennart Åqvist, Brian Chellas, Fredric Fitch, Stig Kanger, Ingmar Pörn, Krister Segerberg, Franz von Kutschera, and G. H. von Wright.[1] In the past few years, one aspect of this tradition has received renewed attention in a series of papers by Nuel Belnap and Michael Perloff, beginning with [7].

Belnap and Perloff describe their approach as *stit theory*, because it concentrates on a construction of the form "α (an agent) sees to it that A," usually abbreviated simply as

$$[\alpha \; \mathit{stit}\colon A].$$

The theory provides a precise and intuitively compelling semantic account of this *stit* operator within an overall logical framework of indeterminism; the account is then used as a springboard for investigating a number of topics from the general logic of agency, such as the proper treatment of certain concepts naturally thought of as involving iterations of the agency operator, as well as interactions of this operator with other truth functional and modal connectives.

The purpose of the present paper is to describe the semantics and explore the applications of an alternative modal agency operator, closely related to that of Belnap and Perloff, but simpler and for certain purposes more natural as an analysis of agency. This alternative operator first appeared, it turns out, in von Kutschera's [43], prior to the work of Belnap and Perloff; it was later suggested independently by John Horty [24], explicitly as an alternative to the account of agency put forth by Belnap and Perloff.

In order to distinguish between the two agency operators under discussion, and for other reasons that will soon become apparent, we describe the original operator presented by

[1] A brief historical sketch of the subject, with references to the works of these writers and others, can be found in Belnap [3]; a more extensive history is contained in Segerberg [36].

Belnap and Perloff as the *achievement* stit, represented here as *astit*; and we describe the alternative suggested by von Kutschera and Horty as the *deliberative* stit, represented here as *dstit*. When we speak simply of a stit operator—or use *stit* alone as a connective in some formula—we mean to generalize over both the deliberative and achievement stit operators, and perhaps others of the same family.

The paper is organized as follows. Section 2 reviews the framework of indeterminism that forms the background for stit theory, and then presents the semantics of both the achievement and deliberative stit operators. Section 3 then compares the deliberative and achievement stits from a more general logical perspective, explores some other logical issues concerning the deliberative stit, and describes a related stit operator. Section 4 studies the concepts of refraining (or omitting) and ability from the perspective of the deliberative stit. And Section 5 explores the use of the deliberative stit operator in the context of deontic logic.

2 The two stits

2.1 Background: branching time

Stit theory is cast against the background of an indeterministic temporal framework—in particular, the theory of branching time due originally to Arthur Prior [34], and developed in more detail by Richmond Thomason in [37] and [40]. The theory 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 a treelike ordering of these moments—an ordering such that, 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 the statement that $m \in h$ can be taken

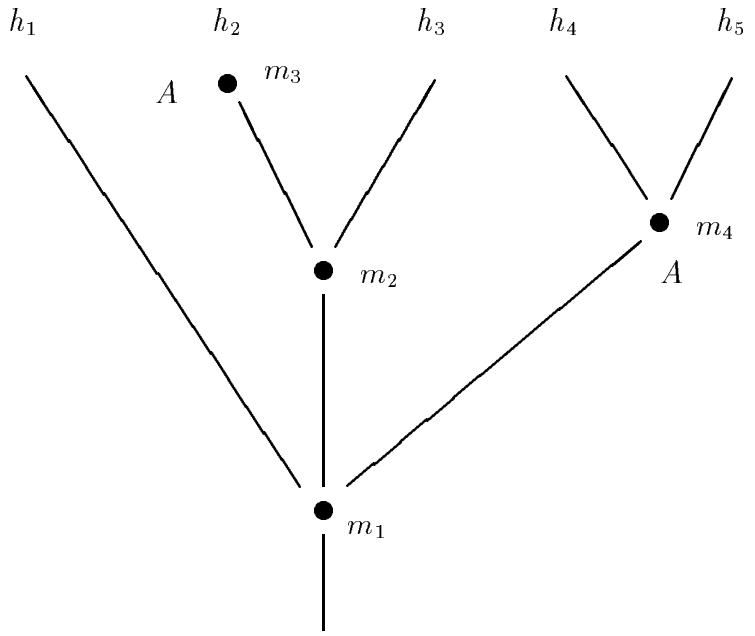

Figure 1: Branching time: moments and histories.

to mean that m occurs at some point in the course of the history h. 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:

- $\mathcal{M}, m/h \models A$ iff $m/h \in v(A)$ for A an atomic formula.[2]

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

- $\mathcal{M}, m/h \models A \wedge B$ iff $\mathcal{M}, m/h \models A$ and $\mathcal{M}, m/h \models B$,

[2]It is not usual for languages of this kind to admit the possibility that even atomic formulas might be true at one index m/h but false at another index m/h', for different histories h and h' belonging to $H_{(m)}$. What we have in mind are situations such as the following. If, in a restaurant, Karl is offered cake or pie for dessert, it seems that "Karl chooses pie," which is at least not obviously non-atomic, might be true relative to one history through m, but false relative to another. In any case, whether or not indexing atomic formulas to both moments and histories is actually necessary for evaluating statements of this kind, allowing for the possibility at least does no harm.

- $\mathcal{M}, m/h \models \neg A$ iff $\mathcal{M}, m/h \not\models A$,
- $\mathcal{M}, m/h \models \mathsf{P}A$ iff there is an $m' \in h$ such that $m' < m$ and $\mathcal{M}, m'/h \models A$,
- $\mathcal{M}, m/h \models \mathsf{F}A$ iff 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* if it is true at every index—in this case, every m/h pair—in every model. 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 the additional concept of settledness, or historical necessity, along with the 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:

- $\mathcal{M}, m/h \models \Box A$ iff $\mathcal{M}, m/h' \models A$ 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 philosophical issues, such as the representation of indeterminism, for which standard tense logic is no help. Details and references can be found in Thomason [40], with an extended discussion of indeterminism in Belnap and Green [6].

2.2 The achievement stit

The stit operator introduced in [7] is designed to approximate the idea of seeing to it that. More exactly, a statement of the form [α *astit*: A] should be taken to mean something like: the present momentary fact that A is guaranteed by a prior choice of the agent α. And it is for this reason, because it is used to describe the present momentary outcome of an agent's prior activity, that we characterize this operator as the *achievement* stit.

In order to capture the meaning of the achievement stit, we must be able to speak of an individual agent's choices, and also, evidently, of the present. As a means of representing these concepts, the basic framework of branching time is supplemented in [7] with three additional primitives.

The first is simply a set *Agent* of agents, individuals thought of as making choices, or acting, in time.[3]

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 acts as instantaneous. In this rarefied environment, the idea of acting or choosing can be thought of simply as constraining the course of events

[3]Other stit papers, while expressly designed to contribute to our understand of agency, and therefore of action, tried to avoid using language that might suggest that the authors understood the ontology of actions. (The relation of stit semantics to some of the previous philosophical work on agency and action is discussed in Perloff [32].) In the present paper, we have been somewhat more relaxed in informal passages about using devices such as singular terms that purport to refer to actions as things in the world, but we should nevertheless be understood in exactly the same spirit. For example, when we say "moment of action," we certainly mean to be calling the reader's attention to a particular moment, but we do not intend to suggest that we understand what, if anything, could be meant by saying that there is an x such that x is an action and x is located at that particular moment. Roughly the same remarks hold for "moment of choice," which we are using as interchangeable with "moment of action" in spite of the following: literary convention easily permits using "moment of choice" for an earlier moment of indecision, while tending to reserve "moment of action" for a later moment shortly after "the action" has commenced. This literary distinction—reminiscent also of Zeno—suggests to us the importance of highlighting the transition from "not-having-acted (or chosen)" to "having-acted (or chosen)." The present paper, however, makes no more of this suggestion.

to lie within some definite subset of the possible histories still available. When Jones butters the toast, for example, the nature of his act, 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 act 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.

The second primitive introduced in [7], then, is a device for representing the constraints that an individual is able to exercise upon the course of history at a given moment, the acts or choices open to him at that moment. Formally, these constraints can be encoded through a choice function, mapping each agent α and moment m into a partition $Choice_\alpha^m$ of the histories $H_{(m)}$ through m. The equivalence classes belonging to $Choice_\alpha^m$ can be thought of as the possible choices or actions available to α at 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 $Choice_\alpha^m$ within which the future course of history must then lie, but that this is the extent of his influence. As additional notation, we let $Choice_\alpha^m(h)$ (defined only when $h \in H_{(m)}$) represent the particular possible choice from $Choice_\alpha^m$ containing the history h. And of course, in order for this choice information to make any sense, we must require that any two histories in $H_{(m)}$ that have not yet divided at m must lie within the same possible choice; the choices available to an agent at m should not allow a distinction between two histories that do not divide until some later moment.

The information represented through these 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 possible choices or actions available to α at those moments. For example, α has three possible choices at m_1—$Choice_\alpha^{m_1} = \{\{h_1, h_2\}, \{h_3\}, \{h_4, h_5, h_6\}\}$—and two at m_2. Because h_1 and h_2 are still undivided at m_1, they must fall within the same partition there, and likewise for h_4 and h_5. At m_3 the agent α 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

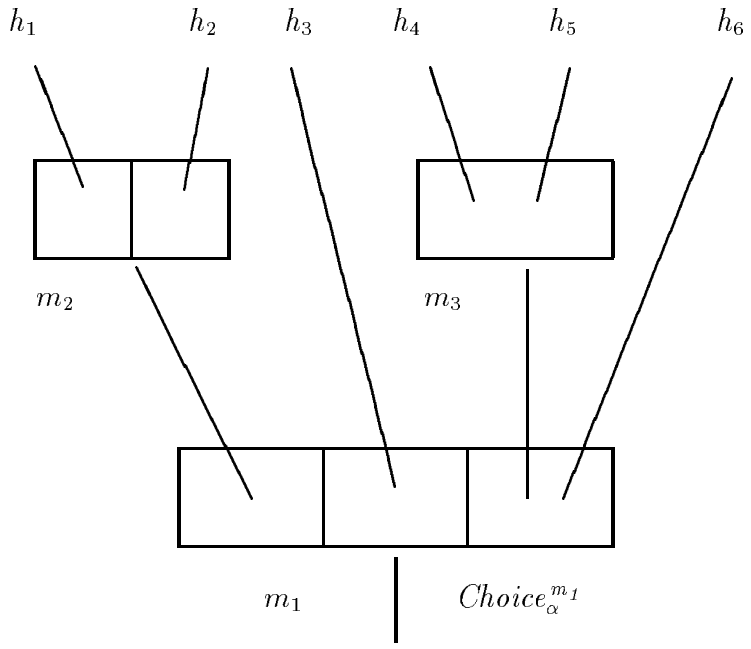

Figure 2: An agent's choices.

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.

The final primitive supplied by [7] is a set *Instant* of instants partitioning the moments of *Tree* horizontally into equivalence classes. Intuitively, an instant represents a set of contemporaneous moments from each of the various histories, with the different moments belonging to a single instant thought of as occurring at the "same time" in the different histories. The instant containing the moment m is represented as $i_{(m)}$. It is supposed that each instant meets (intersects with) each history at exactly one moment, and that instants respect the temporal order of histories in the following sense: if the moment at which an instant i_1 meets a history h is later than the moment at which i_2 meets h, then the same relation holds between the moments at which the instants i_1 and i_2 meet any other histories. These suppositions about instants amount to strong restrictions on the structure of *Tree*, satisfiable only if all histories share an isomorphic temporal ordering, which is then inherited

by the instants themselves. The point of the restrictions, of course, is to allow for temporal comparisons between moments from different histories.

When the basic framework of branching time is supplemented with these additional primitives, the result is a *stit frame* of the form $\langle Tree, <, Agent, Choice,$
$Instant\rangle$, with *Tree* and $<$ as before; and we can define *stit models* as structures of the form $\mathcal{M} = \langle \mathcal{F}, v\rangle$, in which $\mathcal{F}$ is a stit frame and v valuation mapping each propositional constant, as before, into a set of m/h pairs. It is these structures that provide the backdrop for the semantics of the achievement stit; the claim is that the structures are not just mathematical curiosities, but describe—up to a legitimate idealization—the world in which agents act.

Before stating the evaluation rule for the achievement stit, we first require an auxiliary definition. Suppose that the moments m_1 and m_2 occur at the same instant ($i_{(m_1)} = i_{(m_2)}$), and consider some moment w prior to both ($w < m_1$ and $w < m_2$). If m_1 and m_2 lie on histories belonging to the same $Choice_\alpha^w$ partition, these two moments are then said to be $Choice_\alpha^w$*-equivalent.* The idea behind this definition is that, through his choice at w, the agent α can guarantee that whatever moment occurs at the instant $i_{(m_1)}$ $(= i_{(m_2)})$ will lie within some particular $Choice_\alpha^w$-equivalence class, but there is nothing he can do to determine which of the moments within that class it will be.

Using this auxiliary concept, the rule for evaluating an achievement stit at an index m/h of a stit model $\mathcal{M}$ can now be set out as follows:

- $\mathcal{M}, m/h \models [\alpha$ *astit*: $A]$ iff there is a moment $w < m$ such that (1) for all moments m' $Choice_\alpha^w$-equivalent to m, we have $\mathcal{M}, m'/h' \models A$ for all $h' \in H_{(m')}$; and (2) there is some moment $m'' \in i_{(m)}$ such that $w < m''$ and $\mathcal{M}, m''/h'' \not\models A$ for some $h'' \in H_{(m'')}$.

This formidable definition can be grasped more easily by reference to Figure 3, depicting a situation in which $[\alpha$ *astit*: $A]$ is true at m/h, as a result of an action by α at the prior moment w, known as a *witness.*[4] The evaluation rule embodies two requirements, positive and negative, captured by clauses (1) and (2). The positive requirement is that, as a result

[4] A convention for interpreting these figures: when a formula is written next to a moment, it should be taken as settled true at that moment. Thus, for example, the formula A is taken as settled true at the moment m in Figure 3.

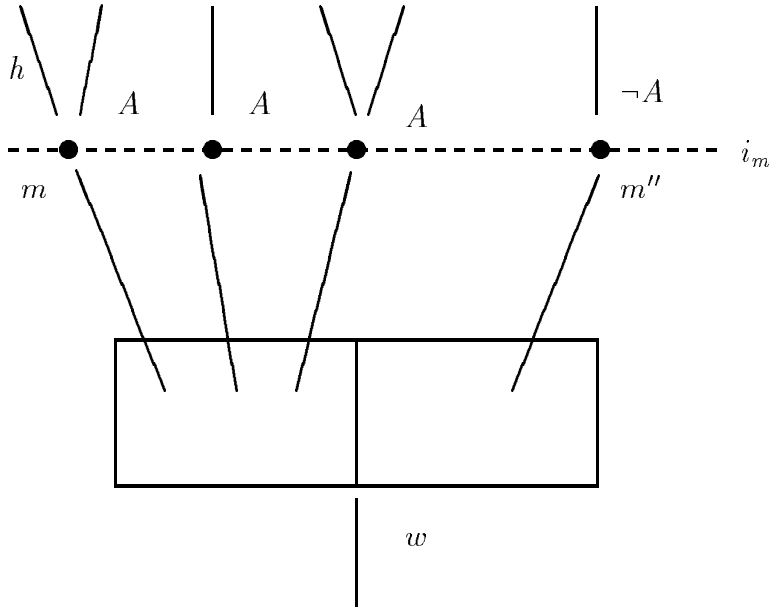

Figure 3: [α *astit*: A] true at m/h.

of a prior choice by α at the witnessing moment w, things have evolved in such a way that A is guaranteed now, at the instant of m, to be true.[5] Of course, since α is unable to determine a single history through his choices at w, he could not have guaranteed that we should now have arrived at m; but he was able to guarantee that we should now have arrived at some moment $Choice_\alpha^w$-equivalent to m, and A is settled true at all of these. The negative requirement is that it was not yet settled at w that A should now be true, so that α's action at w did have some real effect in bringing about the present truth of A.

2.3 The deliberative stit

The semantics of the deliberative stit, like that of the achievement stit, is based on stit frames; accordingly, the two stit operators agree in their underlying view of the relevant structure of the world of agents. The primary conceptual difference between the two operators is this. The truth of an achievement stit [α *astit*: A] depends on two separated moments, the first being the moment at which both the stit formula and the outcome A

[5] In a generalization described most thoroughly in [4], the witness for [α *astit*: A] may also be a chain of moments.

are evaluated, and the second being the required prior moment of choice or action, at which α guarantees the outcome. By contrast, the deliberative stit is referred only to a single moment: a formula of the form [α *dstit*: A] is evaluated at the moment of choice or action, the very moment at which the agent α sees to it that A.

Because it is only the future that can be affected by our actions or choices, it is usually natural to take the complement of a deliberative stit as future tensed; and it is for this reason also that this stit concept is characterized as *deliberative*. The terminology echoes most immediately the notion of deliberative obligation from Thomason [39], but it goes back to Aristotle's observation in the *Nichomachean Ethics* that we can properly be said to deliberate only about "what is future and capable of being otherwise" (1139b7; see also 1112a19-b10).

Since it involves reference only to a single moment/history pair, the truth conditions for the deliberative stit can be stated easily:

- $\mathcal{M}, m/h \models [\alpha \textit{ dstit}: A]$ iff (1) $\mathcal{M}, m/h' \models A$ for each $h' \in \mathit{Choice}^m_\alpha(h)$, and (2) there is some $h'' \in H_{(m)}$ for which $\mathcal{M}, m/h'' \not\models A$.

Evidently, clauses (1) and (2) here are analogous to the positive and negative requirements from the achievement stit. In the present case, the positive requirement is simply that α should act at m in such a way that the the truth of A is guaranteed; α should constrain the histories through m to lie among those on which A is true. The negative requirement, again, is that A should not be settled true, so that α's actions can be seen as having some real effect.

In addition to the primary, one-moment/two-moment contrast between the achievement and deliberative stits, there are two other differences that should be mentioned at once.

The first concerns the role of histories. Although the indices at which an achievement stit is evaluated contain both moments and histories, the histories are present only as a matter of convenience, for reasons of semantic uniformity; they are idle in the evaluation rule. An achievement stit true at some moment/history pair must be true at every history through

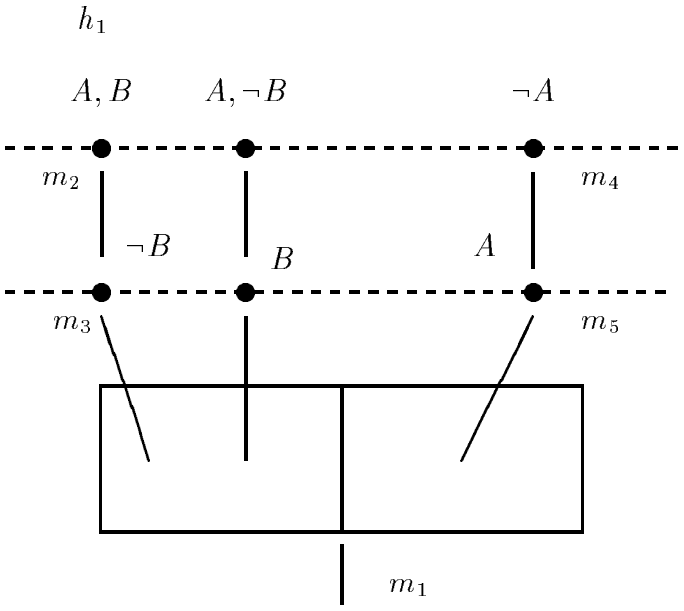

Figure 4: A countermodel to some stit relations.

that moment; this fact is reflected in the validity of

$$[\alpha \; \mathit{astit}\colon A] \supset \Box[\alpha \; \mathit{astit}\colon A],$$

which tells us that any true achievement stit is settled true. For similar reasons, the theory yields also the validity of

$$[\alpha \; \mathit{astit}\colon A] \supset \Box A;$$

the complement of any true achievement stit must itself be settled true.

By contrast, since deliberative stit statements are evaluated at the very moment of an agent's choice or action, histories must play a real role in their evaluation; they provide our only access to the outcome of the agent's action. The theory therefore yields validities such as the following:

$$\neg\Box[\alpha \; \mathit{dstit}\colon A],$$
$$[\alpha \; \mathit{dstit}\colon A] \supset \neg\Box A.$$

The first of these tells us that a deliberative stit is never settled true; the second that it can be true only if its complement is contingent, again reflecting Aristotle's idea that there is no deliberating about matters settled by necessity.

The final point of contrast between the achievement and deliberative stits concerns the role of instants. These play an essential role in the semantics for the achievement stit, but no role at all in the deliberative stit. Because of this, models for evaluating deliberative stits alone can be simpler than the stit models described earlier: they need not contain *Instant* as a primitive, and so do not require us to assume a notion of "same time" across different histories in order to make sense of agency.

One way of understanding the kind of semantic differences between the achievement and deliberative stits that result from the reliance of the former on instants is by considering the following two formulas:

$$[\alpha \ \mathit{astit}\colon A] \supset \mathsf{P}[\alpha \ \mathit{dstit}\colon \mathsf{F}A],$$

$$[\alpha \ \mathit{dstit}\colon \mathsf{F}B] \supset \mathsf{F}[\alpha \ \mathit{astit}\colon B].$$

These formulas may seem to express plausible principles of interaction between the two stits, but in fact, both are invalid, as we can see from the model depicted in Figure 4. Here $[\alpha \ \mathit{astit}\colon A]$ is settled true at the moment m_2, with m_1 as witness: the positive requirement is satisfied because A is true at every moment $\mathit{Choice}_\alpha^{m_1}$-equivalent to m_1, and the negative requirement is satisfied because there is a moment m_4 in $i_{(m_2)}$ at which A is not settled true. However, $[\alpha \ \mathit{dstit}\colon \mathsf{F}A]$ is not true at m_1/h_1: although the positive requirement that $\mathsf{F}A$ should be true at m_1/h' for each h' in $\mathit{Choice}_\alpha^{m_1}(h_1)$ is satisfied, the negative requirement, that there should be some h'' in $H_{(m_1)}$ such that $\mathsf{F}A$ fails at m_1/h'', is not. Is is easy to see also that $[\alpha \ \mathit{dstit}\colon \mathsf{F}B]$ holds at m_1/h_1, but that there is no point in the future of m_1 along h_1 at which $[\alpha \ \mathit{astit}\colon B]$ holds.

3 Some logical points

3.1 Principles of agency

In order to understand the similarities and differences between the deliberative and achievement stit operators from a broader perspective, and also to evaluate the usefulness of each in helping us to understand agency, we now turn to see how some possible theses from the

logic of agency fare under each analysis.[6]

We begin with four principles supported by both stit operators:

$$RE. \quad A \equiv B \ / \ [\alpha \ stit\text{: } A] \equiv [\alpha \ stit\text{: } B],$$

$$C. \quad [\alpha \ stit\text{: } A] \wedge [\alpha \ stit\text{: } B] \supset [\alpha \ stit\text{: } A \wedge B],$$

$$T. \quad [\alpha \ stit\text{: } A] \supset A,$$

$$4. \quad [\alpha \ stit\text{: } A] \supset [\alpha \ stit\text{: } [\alpha \ stit\text{: } A]].$$

The force of the rule RE is that an agent who is responsible for bringing about one state of affairs is likewise responsible for bringing about any logically equivalent state of affairs; this rule seems to make intuitive sense in the present environment, where the intentional components in the concept of action have been set aside. Because of the absence of intentional considerations, the thesis C seems likewise to be justified: one could imagine that an agent might see to it that A holds and that B holds as well without intentionally seeing to it that they hold jointly, but it is hard to deny simply that he does see to it that they hold jointly. And the principle T is again unexceptionable: if an agent sees to it that a certain state of affairs holds, then that state of affairs holds.

The principle 4, on the other hand, does seem to express a substantive claim about agency—that an agent who sees to it that A also sees to it that he sees to it that A—which it is not incoherent to deny; and the principle has been denied in other modal accounts of action.[7] What supports the principle in the present theory is the fundamental assumption that the choices open to an agent at a moment can legitimately be represented as a partitioning of the histories through that moment; and it is this aspect of the underlying framework that must be questioned by anyone who denies 4. Together, of course, the principles 4 and T yield

$$SA. \quad [\alpha \ stit\text{: } [\alpha \ stit\text{: } A]] \equiv [\alpha \ stit\text{: } A],$$

as a modal reduction principle.

[6] Most of the theses in this section have been considered earlier either by Belnap and Perloff or by Chellas [13], and many of the thesis labels are derived from Chellas.

[7] Principles analogous to 4 fail, for example, in the accounts of both Chellas [11] and Brown [10]. Chellas's theory is described later in Section 3.2, and Brown's in Section 4.5.

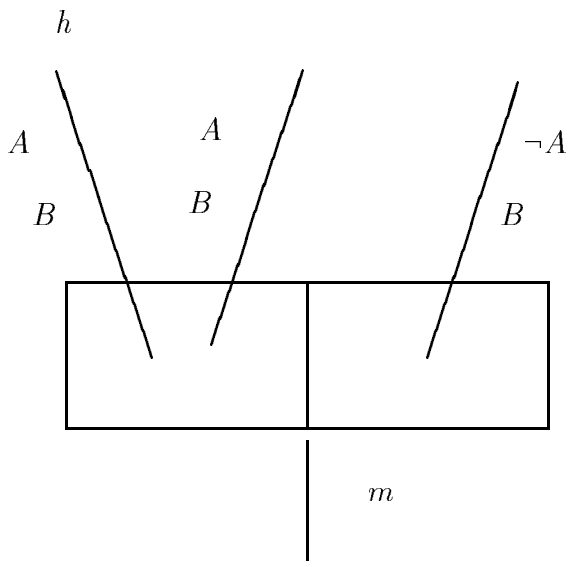

Figure 5: Failure of M for *dstit*.

Among the theses not supported by our two stit concepts, one of the most salient is

$$M. \quad [\alpha \; \textit{stit}\colon A \wedge B] \supset .[\alpha \; \textit{stit}\colon A] \wedge [\alpha \; \textit{stit}\colon B],$$

with is invalid according to both the achievement and deliberative accounts. A countermodel to the achievement stit version of this thesis can be found in [3]; a countermodel to the deliberative stit version appears in Figure 5, where $[\alpha \; \textit{dstit}\colon A \wedge B]$ but not $[\alpha \; \textit{dstit}\colon B]$ is true at m/h.[8] It is, of course, the negative requirement in the evaluation rule for the deliberative stit that is responsible for the failure of M. Although the formula $[\alpha \; \textit{dstit}\colon B]$ must satisfy the positive requirement in this evaluation rule if $[\alpha \; \textit{dstit}\colon A \wedge B]$ does, our example shows that $[\alpha \; \textit{dstit}\colon B]$ need not satisfy the negative requirement even if $[\alpha \; \textit{dstit}\colon A \wedge B]$ does. Because our two stit operators invalidate M, it follows at once that they are not closed under logical consequence; and in fact, although the universally true sentence $\top$ is a logical consequence of any other, both the achievement and deliberative stits actually yield the validity of

$$\overline{N}. \quad \neg[\alpha \; \textit{stit}\colon \top].$$

[8] A further convention for interpreting these 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, for example, the formula A should be taken as true at m/h in Figure 5.

Chellas finds these results—each an upshot of the negative requirement—to be objectionable. Concerning M and closure under consequence, he writes:

> one feels that seeing to a conjunction does imply seeing to the conjuncts and, more generally, that *sees to it that* is closed under consequence. If I see to it that (both) Alphonse is in Alabama and Betty buys a brick, then it follows that I see to it that Alphonse is in Alabama and I see to it that Betty buys a brick. Readers may fashion their own examples and see if they do not concur [13, Section 11].

And concerning $\overline{N}$ and logical truth:

> Can it ever be the case that someone sees to it that something logically true is so? I believe the answer is yes. When one sees to something, one sees to anything that logically follows, including the easiest such things, such as those represented by $\top$. One should think of seeing to it that (e.g.) $0 = 0$ as a sort of trivial pursuit, attendant upon seeing to anything at all [13, Section 12].

We will return later, in Section 4.3, to these objections concerning the negative requirement in an agency operator; we suggest there that, although the need for the negative requirement may not be so clear in simple constructions of the kind that Chellas considers, the advantages of this requirement become more apparent when we focus on certain nested stit constructions. Still, it is worth noting that, even if we restrict ourselves to simple, non-nested stit constructions, intuitions concerning the negative requirement are not uniform: although Chellas feels that closure of the agency operator under logical consequence is intuitively appealing, this view runs against a certain tradition in philosophy, at least. Anthony Kenny, for example, writes as follows on a related issue:

> The President of the United States has the power to destroy Moscow, i.e., to bring it about that Moscow is destroyed; but he does not have the power to bring it about that either Moscow is destroyed or Moscow is not destroyed. The power to bring it about that either p or not p is one which philosophers, with the exception of Descartes, have denied even to God [29, p. 214].

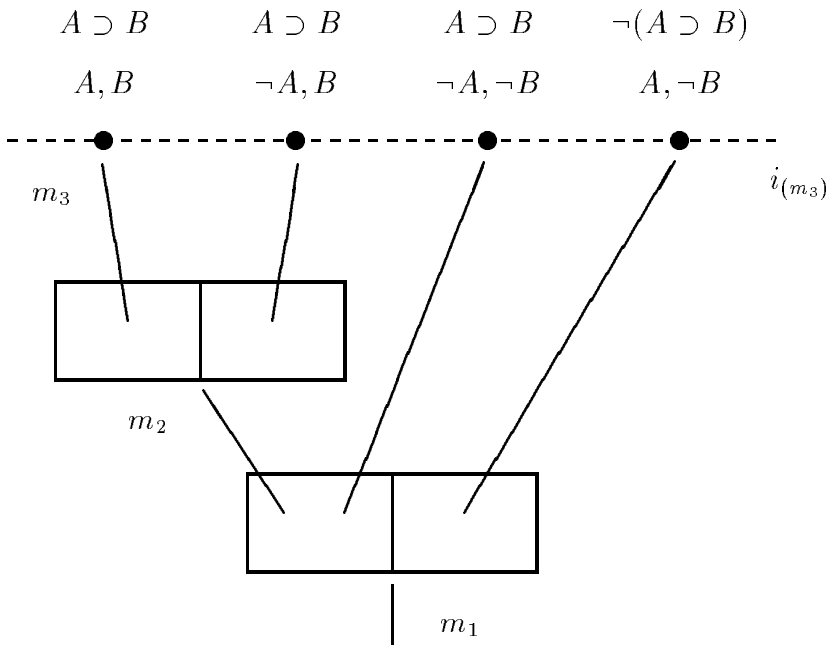

Figure 6: Failure of *SMP* for *astit*.

Finally, we wish to consider an issue closely related to closure under logical consequence: the issue concerning the closure of our two stit operators under modus ponens, or the validity of

$$SMP. \quad [\alpha \; stit\text{: } A] \wedge [\alpha \; stit\text{: } A \supset B] \supset [\alpha \; stit\text{: } B].$$

This thesis is not valid according to the achievement stit. A countermodel is displayed in Figure 6. Here, both $[\alpha \; astit\text{: } A \supset B]$ and $[\alpha \; astit\text{: } A]$ are true at m_3; the moment m_1 is a witness for the first formula, and m_2 a witness for the second. However, $[\alpha \; astit\text{: } B]$ is false at m_3; neither m_1 nor m_2 can witness its truth, since the first of these moments fails to satisfy the positive requirement involved in the achievement stit, and the second fails to satisfy the negative requirement.

This counterexample to *SMP* relies heavily on the temporal distance permitted by the achievement stit between the moment at which an achievement stit formula is evaluated and the required prior moment of the agent's choice—in this case, between m_3 and m_1. The distance leaves room for intervening choices, such as that at m_2. At this intervening moment, B has become inevitable through no agency of α, whereas A remains open to α's

choice. Without this possibility, the counterexample could not be constructed.

Because the deliberative stit is evaluated at the very moment of an agent's choice or action—so that no temporal distance is permitted between action and evaluation—it is easy to see that, in contrast to the achievement stit, this stit concept is closed under modus ponens: it validates *SMP*. The argument is straightforward. Suppose that both [α *dstit*: A] and [α *dstit*: $A \supset B$] hold at the at the pair m/h. By the positive requirement, we know that both A and $A \supset B$ must hold at m/h', for each h' in $Choice^m_\alpha(h)$. Therefore, B also must hold at m/h' for each h' in $Choice^m_\alpha(h)$; and so the positive requirement is satisfied for [α *dstit*: B] to hold at m/h. Again, since [α *dstit*: $A \supset B$] holds at m/h, the negative requirement tells us that $A \supset B$ must fail to hold at m/h'' for some h'' in $H_{(m)}$. From this we can conclude that B fails to hold at m/h''; and so the negative requirement also is satisfied for [α *dstit*: B] to hold at m/h. Since both the positive and negative requirements are satisfied, [α *dstit*: A] must hold at m/h.

3.2 Yet another stit concept

The first analysis of agency developed within the context of modern intensional semantics was set out nearly twenty five years ago, in Chellas's [11].

There are two primary differences between Chellas's theory and the kind of analysis proposed here. First, Chellas's theory is cast against the background of a temporal framework slightly different from the framework of branching time that underlies stit semantics; but it is easy enough to transpose his account to the present environment. More important, Chellas's analysis of agency takes as its fundamental semantic primitive, not a choice partition, but a binary relation—which can be represented in the present environment as a relation R^α_m, (defined for each agent α and moment m) over the histories belonging to $H_{(m)}$. Chellas describes histories standing in the R^α_m relation as "instigative alternatives" for the agent α at the moment m; and he defines his agency operator so that the formula representing "α sees to it that A" is true at an index m/h whenever A is true at m/h' for each instigative alternative h' to h, for each h' such that $R^\alpha_m(h, h')$.

We can approximate Chellas's theory within the current context by linking his notion

of instigative alternativeness to the present idea of choice partitions in the most natural way: stipulating that $R^{\alpha}_{m}(h, h')$ just in case $h' \in \mathit{Choice}^{m}_{\alpha}(h)$, so that histories count as instigative alternatives for α at m whenever they belong to the same cell in the $\mathit{Choice}^{m}_{\alpha}$ partition. In his own theory, Chellas requires only that each R^{α}_{m} relation should be reflexive, but the current approximation yields stronger constraints on these relations of instigative alternativeness: because $\mathit{Choice}^{m}_{\alpha}$ is a partitioning of $H_{(m)}$, the R^{α}_{m} relations defined in the suggested way will turn out to be equivalence relations.

Let us now introduce the operator *cstit*—for "Chellas stit"—as an analog in the context of stit semantics to Chellas's original agency operator. An exact translation of Chellas's own semantics would yield an evaluation rule of the form

- $\mathcal{M}, m/h \models [\alpha\ \mathit{cstit}\colon A]$ iff $\mathcal{M}, m/h' \models A$ for all h' such that $R^{\alpha}_{m}(h, h')$;

but given the current definition of R^{α}_{m}, this is, of course, equivalent to the rule:

- $\mathcal{M}, m/h \models [\alpha\ \mathit{cstit}\colon A]$ iff $\mathcal{M}, m/h' \models A$ for all $h' \in \mathit{Choice}^{m}_{\alpha}(h)$.

Evidently, the evaluation rule for this *cstit* operator is like that for the deliberative stit, except that the negative condition is missing, in keeping with Chellas's reservations concerning this condition. Because of the absence of the negative condition, and because the choice cells partition the histories through a given moment into equivalence classes, it is clear that our defined Chellas stit (unlike Chellas's own agency connective) is an S5 modal operator.

From the following equivalences, it is evident also that in the presence of historical necessity the Chellas and deliberative stits operators are interdefinable:

$$[\alpha\ \mathit{dstit}\colon A] \equiv .[\alpha\ \mathit{cstit}\colon A] \wedge \neg\Box A.$$
$$[\alpha\ \mathit{cstit}\colon A] \equiv .[\alpha\ \mathit{dstit}\colon A] \vee \Box A.$$

Given these straightforward interdefinability relations, the question as to which of these two operators more accurately represents our everyday notion of "seeing to it that" may not be such an important issue; perhaps it would be best to appeal to both operators, for different analytic purposes.[9]

[9]Chellas nevertheless argues that an operator without the negative condition, such as *cstit*, should be

3.3 An axiomatization

A number of issues concerning the axiomatization and decidability of various stit theories, as well as several model theoretic problems, have been studied by Ming Xu, and are discussed in [47].

Here we mention only Xu's axiomatization of a simple case of the theory of the deliberative stit, omitting tense operators and multiple agents, confining attention only to truth functions and stit statements involving a single agent. The axiomatization takes *dstit* and the historical necessity operator $\Box$ as primitive, while including the Chellas stit as a defined connective, with $[\alpha\ \mathit{cstit}\colon A]$ defined as $[\alpha\ \mathit{dstit}\colon A] \vee \Box A$. The basis for the axiomatization is the set of tautologies. We then postulate that each of $\Box$ and *cstit* is an S5 modality:

$$\Box(A \supset B) \supset .\Box A \supset \Box B,$$
$$\Box A \supset A,$$
$$\neg\Box\neg A \supset \Box\neg\Box\neg A,$$
$$[\alpha\ \mathit{cstit}\colon A \supset B] \supset .[\alpha\ \mathit{cstit}\colon A] \supset [\alpha\ \mathit{cstit}\colon B],$$
$$[\alpha\ \mathit{cstit}\colon A] \supset A,$$
$$\neg[\alpha\ \mathit{cstit}\colon \neg A] \supset [\alpha\ \mathit{cstit}\colon \neg[\alpha\ \mathit{cstit}\colon \neg A]].$$

---

regarded as fundamental:

> It may be that conversational assertions of agency using "sees to it that" carry an implication of seeing to it *really*, but even if so this is no license for making a negative stipulation intrinsic to the meaning of this idiom. To argue for the necessity of a negative half to the truth conditions for [a stit operator] one must first demonstrate that there is no adequate account of *sees to it–really* in which such negativity is external. For example, one might investigate the meaning of $[[\alpha\ \mathit{cstit}\colon A] \wedge \neg\Box A] \ldots$ [13, Section 15].

To the extent that we understand this argument, we think we disagree. It seems to us that it is consistent both to recognize the fact that the deliberative stit is definable through the Chellas stit together with an "external" statement of the negative condition and to suggest that this more complicated, defined concept is actually the one that corresponds more closely to our ordinary notion. In any case, since the Chellas stit can also be defined by disjoining the deliberative stit with an external statement of non-negativity, the situation appears to be symmetric.

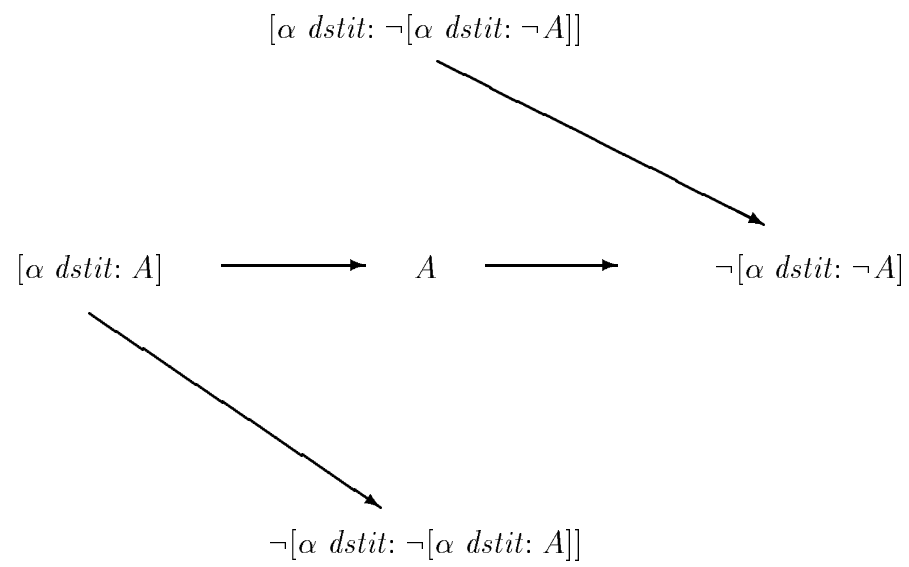

Figure 7: Positive *dstit* modalities.

And a final axiom relates the two primitives in the obvious way:

$$[\alpha \; dstit\text{: } A] \supset \neg\Box A.$$

The rules are *modus ponens* and the rule RN: from A to infer $\Box A$.

3.4 Deliberative stit modalities

A *stit modality* can be defined as any sequence of zero or more negations and stit operators affixed to some schematic argument. Such a modality can be defined as *positive* or *negative* according to whether it contains an even or odd number of negations; and any two such modalities can be defined as *equivalent* if they imply each other (whenever they are affixed to the same formula as argument), and *distinct* otherwise.

The theory of the deliberative stit allows for ten distinct modalities, with the five positive ones organized as in Figure 7. The picture for the negative deliberative stit modalities can be obtained by affixing an initial negation to each formula in this figure, and then reversing the arrows. Ignoring both A and $\neg A$, we can thus see that the deliberative stit presents us with four "modes of action" and four "modes of inaction," organized just so.

It is of historical interest to note that the alethic modal system S4.2 defined by Michael

Dummett and E. J. Lemmon [17] contains exactly the the same ten modalities as the deliberative stit, but with a different structure. In addition to the implications set out in Figure 7, the Dummett-Lemmon system also allows implications among the positive modalities (and their negative analogs) that would be analogous to our

$$\neg[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]] \supset [\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon \neg A]],$$
$$\neg[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]] \supset \neg[\alpha \; \mathit{dstit}\colon \neg A],$$
$$[\alpha \; \mathit{dstit}\colon A] \supset [\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon \neg A]].$$

But it is easy to verify that each of these fails for the deliberative stit.

4 Refraining and ability

4.1 Refraining: basic analysis

The concept of refraining, or omitting, was characterized by von Wright [44, p. 45] as a "correlative" of action; but even among philosophers explicitly concerned with action, this correlative notion is seldom treated in any detail, perhaps because it is so difficult to understand. When an agent refrains from smoking, for example, he does not smoke; but there seems to be more to it than that. An agent is not naturally thought to refrain from doing whatever it is he does not do—particularly, as von Wright notes, when those actions lie beyond his capacity. Even if it is true that some agent does not alter the course of a tornado, for example, it still does not seem correct to say that he refrained from doing so.

Because refraining involves more than simple not doing, some writers have pursued the strategy of conjunctive definition, attempting to characterize refraining as not doing plus "something else." One example is von Wright himself, who feels that the concept of refraining cannot be defined in terms of action and truth functional connectives alone, at least if action is to be analyzed as he proposes. Instead, he suggests that refraining should be defined as not acting conjoined with the ability to act: an agent refrains from doing a certain thing if and only if "he *can do* this thing, but *does* in fact *not do* it" [44, p. 45].[10]

[10]The mode of action described here as refraining is characterized by von Wright in [44] as "forbearing" and in [46] as "omitting." Von Wright notes that refraining (forbearing, omitting), as analyzed in his work, is

Working from the perspective of the achievement stit, Belnap and Perloff rejected this suggestion of von Wright's in [7, Section 5.1]. Instead, they choose to develop another theme also present in von Wright—that refraining, although it involves not doing, is itself a kind of doing, a "mode of action or conduct" [46, p. 12]. When an agent refrains from smoking, he does not smoke; but not smoking itself seems to be something he does. In the context of stit semantics, not smoking is represented as not seeing to it that one smokes; and so it seems that refraining from smoking—performing the action of not smoking—can be represented as seeing to it that one does not see to it that one smokes. More generally, it is suggested in [7] that the idea that α refrains from seeing to it that A can be represented through a stit statement of the form

$$[\alpha \; \mathit{stit}\colon \neg[\alpha \; \mathit{stit}\colon A]].$$

Evidently, this analysis casts refraining as a concept definable in terms of an agency operator and truth functions alone, contrary to von Wright's view that additional linguistic resources are necessary; and the source of this difference is easy to see. Unlike von Wright's representation of action, which forbids nesting, stit operators do allow the nesting of one action expression within another; and this ability to nest is crucial for the analysis of refraining proposed above.

4.2 Refraining from refraining

Since stit operators encourage nesting, inviting us to define concepts such as refraining, it is natural to consider also more complicated, deeply nested concepts, such as refraining from refraining. Given our current analysis of refraining, the idea that α refrains from seeing to it that he refrains from seeing to it that A translates into the formula

$$[\alpha \; \mathit{stit}\colon \neg[\alpha \; \mathit{stit}\colon [\alpha \; \mathit{stit}\colon \neg[\alpha \; \mathit{stit}\colon A]]]],$$

the logically weakest member in a series of progressively stronger notions. Refraining from an action involves the ability to perform that action, but not necessarily any awareness of that ability; stronger notions can be obtained if one requires an awareness of the ability, an actual decision to refrain, or a decision to refrain in the face of inclination (which he describes as "abstaining").

which says, of course, that α sees to it that he does not see to it that he sees to it that he does not see to it that A. This is equivalent by the principles SA and RE to the marginally less confusing

$$[\alpha \; \mathit{stit}\colon \neg[\alpha \; \mathit{stit}\colon \neg[\alpha \; \mathit{stit}\colon A]]],$$

telling us that α sees to it that he does not see to it that he does not see to it that A.

We can now turn to a question considered by Meinong in the manuscript of his *Ethische Bausteine*:

> One may ask whether the essential features of the law of omission are to be found in the law of double negation or in some analogues thereof. In such a case omission of omission would yield commission, just as the negation of a negation yields an affirmation ....[11]

In the present context, this question—whether refraining from refraining is equivalent to doing—can be cast as a question concerning the validity of the formula

$$RR. \quad [\alpha \; \mathit{stit}\colon A] \equiv [\alpha \; \mathit{stit}\colon \neg[\alpha \; \mathit{stit}\colon \neg[\alpha \; \mathit{stit}\colon A]]],$$

dubbed in [4] as the "Refref conjecture." And this again, is a matter on which the achievement and deliberative stits differ.

Given only the definition of the achievement stit, RR can be shown to be invalid, since there are mathematically correct stit models in which it is false. However, as detailed in [3], the situation is more complicated than this simple statement suggests. It turns out that the only models in which RR can be falsified are those involving agents, known as *busy choosers*, who make an infinite number of nonvacuous choices between two given moments. If there are no busy choosers in reality, then RR is valid for the achievement stit in the more restricted class of real models.[12]

[11] This passage is cited in footnote 15 of Chisholm [14], which is where we learned of Meinong's concern with the issue.

[12] We note also that if there are no busy choosers, so that RR is validated, then the structure of the achievement stit modalities is exactly as described for the deliberative stit in Section 3.4.

In the case of the deliberative stit, the situation is less complicated: here, RR is simply valid—the deliberative stit tells us that refraining from refraining is equivalent to doing. We omit the straightforward proof.

It is worth noting that RR is valid also according to the Chellas stit operator, since of course *cstit* is an S5 modality.

4.3 Can do otherwise

It is interesting that the question whether refraining from refraining is equivalent to doing can be formulated so clearly in the framework set out here; but the question itself is perhaps not terribly important: it is hard to think of any fundamental philosophical views that would be shattered either by a positive or a negative answer. A more important issue concerns the proper sense, if any, in which performing an action can be said to imply the ability to do otherwise. That acting does imply the ability to do otherwise is a view going back, again, at least to Aristotle, who writes in the *Nichomachean Ethics* that "where it is in our power to do something, it is also in our power not to do it, and when the 'no' is in our power, the 'yes' is also" (1113b7-8); and the topic has been much debated in the contemporary literature as well.[13]

Within the context of stit semantics, the idea that performing an action implies the ability to do otherwise has been considered earlier, in [8]. For the most part, that paper concentrated on the achievement stit, and used this connective to provide clear formulations of a number of senses both of what it could mean to "do otherwise" and of how an agent's "ability" should be understood.

In the present paper, we concentrate instead on the deliberative stit; and rather than consider a variety of interpretations, we focus only on a single way of understanding the idea that acting implies the ability to do otherwise. First, we assume that "doing otherwise" is refraining; when acting is represented as seeing to it that some proposition holds, the ability to do otherwise is to be represented as the ability to refrain from seeing to it that the proposition holds. Second, we assume quite generally that an agent's ability (personal

[13]See, for example, Chisholm [16], Frankfurt [20], and van Inwagen [42].

can-do) can be represented through a simple combination of ordinary historical possibility (impersonal can) and the deliberative stit (personal to-do), so that the formula

$$\Diamond[\alpha \; \mathit{dstit}\colon A]$$

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

Together, these two assumptions allow us to express the idea that acting implies the ability to do otherwise as the principle

$$\mathit{ACR}. \quad [\alpha \; \mathit{dstit}\colon A] \supset \Diamond[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]];$$

and it is a simple matter to see that this principle is valid. Suppose $[\alpha \; \mathit{dstit}\colon A]$ holds at m/h. By the negative requirement in the deliberative stit evaluation rule, we know that there must be some h' such that A is false at m/h'. It is then easy to see that $\neg[\alpha \; \mathit{dstit}\colon A]$ holds at m/h'' for each h'' in $\mathit{Choice}^m_\alpha(h')$. The positive condition is thus satisfied for $[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]]$ to hold at m/h', and the negative condition is satisfied also, since $[\alpha \; \mathit{dstit}\colon A]$ is true at m/h by assumption. Therefore, $[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]]$ is true at m/h'; and so $\Diamond[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]]$ must be true at the original index m/h.

Not only does acting imply the ability to refrain, according to this analysis, but it turns out also, as it should, that refraining from acting entails the ability to act; the principle

$$\mathit{RCA}. \quad [\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]] \supset \Diamond[\alpha \; \mathit{dstit}\colon A]$$

is likewise valid. It would be easy enough to supply a semantic argument, as above, for the validity of this principle; but there is no need, since the principle follows from others already established. Suppose α refrains from seeing to it that A:

$$[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]].$$

This action of refraining is, of course, itself a doing, and so itself, according to ACR, something that the agent should be able to refrain from:

$$\Diamond[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]]].$$

But now, this formula attributes to α the ability to refrain from refraining from seeing to it that A; and since the deliberative stit satisfies the principle RR, telling us that refraining from refraining is equivalent to acting, we can conclude:

$$\Diamond[\alpha \; \mathit{dstit}\colon A].$$

As a final observation concerning the relations sanctioned by the deliberative stit among acting, refraining, and ability, we note that, since both ACR and RCA are valid, and since historical necessity is an S5 modality, ordinary modal reasoning allows us to conclude:

$$\mathit{CACR}. \quad \Diamond[\alpha \; \mathit{dstit}\colon A] \equiv \Diamond[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]].$$

This formula can be taken as expressing the Aristotelian principle, cited above, that the ability to act coincides with the ability to refrain from acting. Such a principle of two-way ability is advanced by Kenny, for example, who argues that it can be used to distinguish "full-blooded" abilities, for whose exercise we can be held responsible, from mere "natural powers," such as the power to grow old [29, pp. 226–228].[14] And the principle has been endorsed also by von Wright as a "fundamental law of ability logic," according to which "ability to do and to omit [refrain] are reciprocal" [45, p. 391].

We feel that the validity of ACR, RCA, and the Aristotelian principle $CACR$ help to support both the schematic analysis of refraining in terms of a nested stit operator, and also the helpfulness of the deliberative stit in this analysis. It is important to note, however, that all three of these principles depend on the presence of the negative requirement in the semantics of the deliberative stit: if the deliberative stit were replaced with the Chellas stit, from which the negative requirement is absent, none of these principles would be valid. The reason for this is that the principles rely upon the distinction between true refraining and simple not doing, and at least given the current analysis of refraining, the negative requirement is crucial for drawing this distinction. The Chellas stit operator does not allow us to distinguish refraining from simple not doing due to the validity of

$$\neg[\alpha \; \mathit{cstit}\colon A] \equiv [\alpha \; \mathit{cstit}\colon \neg[\alpha \; \mathit{cstit}\colon A]].$$

[14] See Kenny [30, pp. 7–9] for a discussion of this distinction in Aristotle.

Earlier, in Section 3.1, we considered the issue as to whether a negative requirement should be included in the semantics an an agency operator. We noted that intuitions concerning the desirability of this requirement are divided when one focuses only on simple, non-nested agency constructions. The negative requirement does receive at least indirect support, however, through its role in allowing us to define a notion of refraining—a nested agency construction—with attractive connections to action and ability. One might feel that its role in validating principles such as *ACR*, *RCA*, and *CACR* provides sufficient reason for accepting the negative requirement, in spite of the divided intuitions concerning its desirability in non-nested constructions.

4.4 Two views of refraining

Let us return to the contrast described in Section 4.1 between the stit analysis of refraining and the earlier proposal of von Wright. The idea that α refrains from seeing to it that A is represented in the current framework—with an agency operator that allows embedding—through a statement of the form

$$[\alpha \; \mathit{stit}\colon \neg[\alpha \; \mathit{stit}\colon A]].$$

Because his treatment of agency does not allow the embedding of one agentive context inside another, this kind of analysis is not available to von Wright; and in fact, he argues that refraining cannot be defined in terms of agency or action alone. Instead, he introduces the concept of ability as a separate primitive notion, and suggests that the idea that α refrains from seeing to it that A should be analyzed as meaning not only that α does not see to it that A, but also that the truth of A is something that α has the ability to bring about. Now, as we have seen, the notion that α is able to see to it that A can be approximated through the stit formula $\Diamond[\alpha \; \mathit{stit}\colon A]$; and so von Wright's analysis of α's refraining from seeing to it that A can itself be approximated through a statement of the form:

$$\neg[\alpha \; \mathit{stit}\colon A] \wedge \Diamond[\alpha \; \mathit{stit}\colon A].^{15}$$

[15]This is not necessarily an approximation that von Wright would accept, since he rejects the idea that "the notion of 'can do' involves a superposition of operators, one for 'can' and another for 'do' ...," and

As we mentioned earlier, Belnap and Perloff reject von Wright's analysis in their [7], the paper that introduces the achievement stit. Things are different from the perspective of the deliberative stit, however. Here it turns out that these two analyses of refraining—the stit analysis and von Wright's—actually coincide; for it is easy to see that the formula

$$[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]] \equiv .\neg[\alpha \; \mathit{dstit}\colon A] \wedge \Diamond[\alpha \; \mathit{dstit}\colon A]$$

is valid. The implication from left to right follows at once from the principle T together with the fact, noted in Section 2.3, that $[\alpha \; \mathit{dstit}\colon A]$ implies $\neg\Box A$ for any statement A. To see the implication from right to left, suppose that $\neg[\alpha \; \mathit{dstit}\colon A] \wedge \Diamond[\alpha \; \mathit{dstit}\colon A]$ is true at m/h. Since $\neg[\alpha \; \mathit{dstit}\colon A]$ holds at m/h, we must have $\neg[\alpha \; \mathit{dstit}\colon A]$ true also at m/h' for each h' in $\mathit{Choice}^m_\alpha(h)$; and so the positive condition is satisfied for $[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]]$ to hold at m/h. But since $\Diamond[\alpha \; \mathit{dstit}\colon A]$ holds also at m/h, there must be some h'' in $H(m)$ such that $[\alpha \; \mathit{dstit}\colon A]$ holds at m/h''. Therefore, the negative requirement is satisfied as well; and so $[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]]$ holds at m/h.

We find this equivalence between the stit analysis of refraining (in its deliberative stit version) and the earlier proposal of von Wright's to be reassuring: it is always nice, and a source of mutual support, when independently motivated proposals for analyzing some phenomenon happen to coincide.

It is interesting to note also that there is a sense in which what von Wright achieves by supplementing his action language with an additional primitive concept of ability is accomplished already in the present framework simply through the presence of nested agency constructions. As we can see from the validity of the formula

$$\Diamond[\alpha \; \mathit{dstit}\colon A] \equiv .[\alpha \; \mathit{dstit}\colon A] \vee [\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]],$$

once nesting is allowed, at least our current approximation of ability can be defined in terms of agency alone.

---

prefers instead to take the notion of ability as primitive [45, p. 391].

4.5 Ability

In our treatment of the ability to refrain from action, and also in our discussion of von Wright's analysis of refraining, we have suggested the formula $\Diamond[\alpha\ \mathit{dstit}\colon A]$ as an approximation of the idea that the agent α has the ability to see to it that A. We now consider the suggestion independently—apart from the connection with refraining—by relating it to two points in the literature.

Kenny's objection

It is a well known thesis of Kenny's, developed in [28] and [29], that the logic of ability cannot be formalized using the techniques of modal logic. Following von Wright in describing the 'can' of ability as a dynamic modality, Kenny puts the point by writing that "ability is not any kind of possibility; ... dynamic modality is not a modality" [29, 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

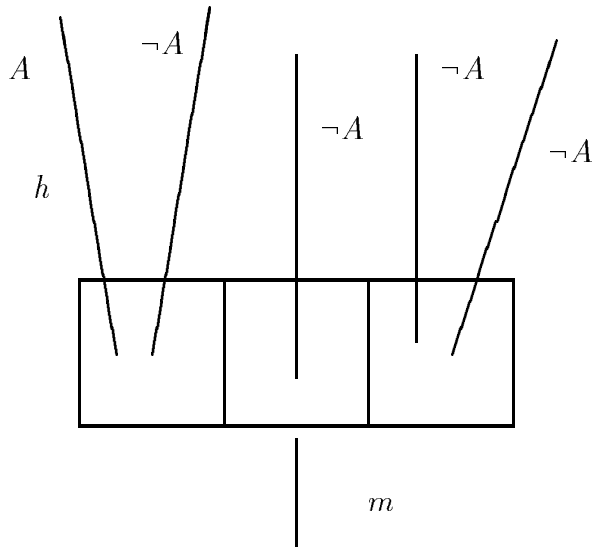

Figure 8: A without $\Diamond[\alpha\ \mathit{dstit}\colon A]$.

$C\Diamond$. However, it is not this possibility operator alone that is taken to represent ability, but rather a combination of historical possibility and the deliberative stit; and the combination fails to satisfy the analogous schemata: both

$$A \supset \Diamond[\alpha\ \mathit{dstit}\colon A],$$
$$\Diamond[\alpha\ \mathit{dstit}\colon A \vee B] \supset .\Diamond[\alpha\ \mathit{dstit}\colon A] \vee \Diamond[\alpha\ \mathit{dstit}\colon B],$$

are invalid. We provide a countermodel only to the first, based on Kenny's darts example, and depicted in Figure 8. Here, m is the moment at which α throws the dart; the cells belonging to Choice^m_α 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, then the dart will hit the bull's eye, but this is not an outcome that the player has the ability to guarantee: although A is true at m/h, the formula $\Diamond[\alpha\ \mathit{dstit}\colon A]$ is not.[16]

[16]It is interesting to note that Kenny himself briefly explores in [29, pp. 226–229] the strategy developed here of representing ability by combining ordinary modal possibility with a special operator representing action. However, the action operator he uses is von Wright's; and Kenny then abandons this strategy for representing ability because of certain problems that he runs into in attempting to develop the idea with this particular operator. The present proposal can thus be seen as developing this idea of Kenny with a different representation of action, the deliberative stit.

Brown's theory

Another response to Kenny's objections against the modal analysis of ability is set out in a paper by Mark Brown [9]. Kenny himself observes that, because the modal schema $T\Diamond$ can be falsified in standard models of modal logics (those based on a binary accessibility relation among possible worlds), the fact that counterexamples to this schema can be constructed using the 'can' of ability does not count as a conclusive argument against the possibility of a modal analysis of this concept. But since $C\Diamond$ is valid in standard models, he judges that the counterexamples to this schema do show that the techniques of possible worlds semantics cannot be used in analyzing the logic of ability.

Brown points out, however, that even this conclusion is too strong, since Kenny limits his attention only to standard models for modal logics, and does not consider more general, non-standard models in which even $C\Diamond$ can be falsified. He then goes on himself to develop an account of ability as a modal operator definable using minimal models—those non-standard models in which accessibility is thought of as relating each individual world not simply to a set of worlds, but to a set of sets of worlds, or propositions.[17]

More exactly, Brown's analysis is based on models of the form $\mathcal{M} = \langle W, N, v\rangle$, in which W is a set of worlds, v is an ordinary valuation, and N is a function mapping each individual world w into some subset $N(w)$ of $\mathcal{P}(W)$, the power set of worlds. Intuitively, the various members of $N(w)$, each a proposition or collection of worlds, represent the results of performing the various actions open to some agent in the world w; and of course, the reason actions are supposed to lead to sets of worlds, rather than individual worlds, is that no agent can determine through his actions every detail of the the resulting situation. The basic idea underlying Brown's analysis is that the agent can be thought of as having the ability, in some world, to bring it about that a proposition A holds just in case there is an action, or choice, open to him in that world whose performance would guarantee the truth of A. If we take $\langle[]\rangle$ as a special modal operator representing the 'can' of ability, this idea then gives rise to the following evaluation rule:

[17] A general treatment of minimal models for modal logic can be found in Chellas [12, Chapter 7].

- $\mathcal{M}, w \models \langle[]\rangle A$ iff there is some action $K \in N(w)$ such that $\mathcal{M}, w' \models A$ for all $w' \in K$.

This operator of Brown's escapes Kenny's objections, allowing the analogs of both $T\Diamond$ and $C\Diamond$ to be falsified; and Brown advances other arguments as well for regarding it as an appropriate formalization for the 'can' of ability. Rather than considering the proposal more closely, however, we simply show here that, in spite of some differences in detail, it is actually quite close in conception to our own suggested analysis.

In order to see this, let us introduce a new, temporary operator *bstit*—for "Brown stit." This operator is supposed to function in the present environment as an analog to Brown's representation of the 'can' of ability, so that $[\alpha\ \mathit{bstit}\colon A]$ means that α has the ability to see to it that A; and in interpreting the operator formally, let us see how Brown's ideas might be adapted from their original minimal model environment to the new context of stit semantics.

One difference between the two contexts is that both agents and temporal information are treated more explicitly in stit models, but Brown himself says that the idea of ability analyzed in his logic is to be construed "neither timelessly nor impersonally," but simply that these matters are left tacit in his approach [9, p. 6]. A second difference is that, in the context of stit models, it is natural to represent the actions or choices available to an agent at a moment, not as sets of moments, but rather as sets of histories through that moment; and it is most natural to use the choice primitive already present in stit models for that purpose, thinking of the possible actions available to the agent α at the moment m as the members of the partition Choice_α^m.

Already, then, we can mirror Brown's analysis in the stit framework, simply by taking $[\alpha\ \mathit{bstit}\colon A]$ as settled true at a moment m whenever there is some possible action or choice K in the partition Choice_α^m such that A is true at the index m/h' for each history h' belonging to K. Such an analysis, however—like Brown's—would attribute to an agent the ability to bring about the truth of any logical consequence of any proposition whose truth he could bring about, and so the ability to bring about the logical truths; and this conclusion runs counter to the viewpoint developed here. In order to bring the suggestion in line with the present point of view, it is necessary only to supplement the analysis with a negative

requirement, so that the agent can be said to have the ability to guarantee the truth of some statement only if its truth is not inevitable. This leads to the evaluation rule:

- $\mathcal{M}, m/h \models [\alpha \ \mathit{bstit}: \ A]$ iff (1) there is a possible action $K \in \mathit{Choice}^m_\alpha$ such that $\mathcal{M}, m/h' \models A$ for each $h' \in K$, and (2) there is some $h'' \in H_{(m)}$ for which $\mathcal{M}, m/h'' \not\models A$.

We hope that the relations are clear between Brown's analysis of ability and the *bstit* operator defined here; our operator is arrived at by, first, adapting Brown's ideas in a straightforward way from their original minimal model environment to the new context of stit models, and then supplementing the result with a negative requirement that is absent from Brown's own analysis. Moreover, the introduction of this *bstit* operator, with its connections to Brown's analysis, allows us also to see the connections between Brown's proposal and our current suggestion of treating ability through a combination of historical possibility and the deliberative stit; for it now turns out that

$$[\alpha \ \mathit{bstit}\colon A] \equiv \Diamond[\alpha \ \mathit{dstit}\colon A]$$

is valid. The ideas underlying the *bstit* operator, with their roots in Brown's work, coincide with those underlying our current suggestion.

Still, it would be a mistake to overemphasize the similarities between Brown's minimal model analysis of ability and our current suggestion, developed in the framework of stit models. Even apart from the more explicit treatment of temporal matters in stit models, and even apart from the negative requirement in our suggestion, there are other important differences between the logics of ability resulting from Brown's analysis and that proposed here. The reason for this is that Brown's minimal models are much less constrained than stit models. Apart from nonemptyness, Brown imposes no conditions at all on the actions or choices open to an agent at a world w, the propositions belonging to $N(w)$. These propositions are not required to exhaust the space of possibilities, so that each world must belong to some member of $N(w)$; and they are permitted to overlap, so that the same world might belong to two different members of $N(w)$. In stit models, however, because the possible actions open to an agent α at a moment m are identified with the members of

$Choice^m_\alpha$, these actions are actually required to partition the relevant set of possibilities, the set of histories belonging to $H_{(m)}$.

Because it places more restriction on the structure of actions, our current suggestion results in a stronger logic of ability than Brown's, validating statements whose analogs in Brown's framework are invalid. For example, Brown's theory allows countermodels to the formula

$$\langle[]\rangle\langle[]\rangle A \supset \langle[]\rangle A,$$

while the analogous statement

$$\Diamond[\alpha \; \textit{dstit}\colon \Diamond[\alpha \; \textit{dstit}\colon A]] \supset \Diamond[\alpha \; \textit{dstit}\colon A]$$

is valid in stit models.

In fact, Brown sees it as an advantage of his account that it does not validate this principle; he views it as an incorrect principle for reasoning about ability, illustrated by the following example:

> Suppose I am a skillful enough golfer that on the short par 3 hole I can hit the green in one stroke, and that, no matter where on the green the ball lands, I can then putt out in one additional stroke. Nonetheless, until I know where the ball lands on the green I don't know which further action to take to get the ball into the hole. It may not be true that I am able to get a hole in one, nor even that there is some pair of strokes I can choose in advance that will assure the ball's going into the hole [9, p. 20].

Apparently, the point of this example is that, at the tee, the golfer is able to get himself into a position from which he will then be able to put the ball into the hole, but that it is incorrect to say of him at the tee simply that he is able to put the ball into the hole. Although there may be a sense in which it can be said that an agent is able to bring about an outcome whenever there is a sequence of actions he can perform that will guarantee its occurrence, we agree with Brown that, at least on the momentary reading of ability, it is incorrect to say of the golfer at the tee that he has the ability then to put the ball into the

hole. Still, this does not necessarily cast doubt on the principle in question; for it is not clear that, in the same momentary sense of ability, the golfer at the tee is able to be able to put the ball in the hole: instead, it seems that what the golfer at the tee is able to do in the momentary sense is to bring it about that in the future he will be able to put the ball in the hole. If this is right, then Brown's example does not undermine the principle stated above, but only the principle

$$\Diamond[\alpha \; \mathit{dstit}\colon \mathsf{F}\Diamond[\alpha \; \mathit{dstit}\colon A]] \supset \Diamond[\alpha \; \mathit{dstit}\colon A],$$

which is indeed invalid in stit models.

5 Oughts and obligations

5.1 Oughts in branching time

Stit theory is informed by the Restricted Complement Thesis [3, p. 787]—the idea that deontic operators (among others) should be restricted to take only stit sentences as their complements. In the present section, however, we relax this constraint, exploring the headway to be gained by allowing the deliberative stit to interact with a more generally applicable deontic operator $\bigcirc$, representing 'It ought to be that ...', which enables us to construct sentences of the form $\bigcirc A$ regardless of the grammatical form of A.

We begin by considering a technique, first set out in Thomason [39], for incorporating a standard deontic operator of this kind into the framework of branching time.[18] As in ordinary deontic logic, the ought operator depends for its interpretation on a nonempty background set of ideal possibilities, those in which things turn out as they ought to; and a sentence $\bigcirc A$ is thought of as true whenever A holds in all of these ideal possibilities. In the context of branching time, the possibilities are realized as histories; moreover, the ideal possibilities at a moment m are limited to a subset of $H_{(m)}$, the histories still available at

[18] Work along similar lines, but against the background of a slightly different temporal framework, had previously been carried out by Brian Chellas [11], Richard Montague [31], and Dana Scott [35]; historical details can be found in Thomason [40].

m. A sentence of the form $\bigcirc A$ is then taken 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 stit frames with a function *Ought* mapping each moment m into a nonempty subset $\mathit{Ought}\,(m)$ of $H_{(m)}$; the result is a structure of the form $\langle \mathit{Tree}, <, \mathit{Agent}, \mathit{Choice}, \mathit{Ought}, \mathit{Instant}\rangle$ with *Tree*, $<$, *Agent*, *Choice*, and *Instant* as before (and in which, of course, *Instant* is not really necessary for the interpretation of the deliberative stit). Where $\mathcal{M}$ is a model that results from interpreting our background language against a structure of this form, the evaluation rule for ought statements is set out as follows:

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

Several logical features of this historical ought should now be apparent. Because $\mathit{Ought}\,(m)$ is a subset of $H_{(m)}$, and a nonempty subset, the ought operator lies between historical necessity and historical possibility: both $\Box A \supset \bigcirc A$ and $\bigcirc A \supset \Diamond A$ are valid. Because $\mathit{Ought}\,(m)$ is, again, nonempty, the theory rules out normative dilemmas, in the sense that $\bigcirc A \wedge \bigcirc \neg A$ is unsatisfiable; and in fact, it is easy to see that this historical ought is a normal modal operator. Finally, statements of the form $\bigcirc A$, again like statements of the form $\Box A$, are always either settled true or settled false.

As a convention in our diagrams, we indicate the histories belonging to $\mathit{Ought}\,(m)$—the ideal histories at m—by marking them with asterisks. Thus, Figure 9, for instance, depicts a situation in which the histories h_1 and h_3 are classified as ideal at the moment m. As a result, we can see that the statement $\bigcirc A$ is settled true at m, while $\bigcirc B$ is settled false.

This is really all that is necessary to know about the historical ought in order to understand its interactions with the deliberative stit (and it is now possible to skip ahead to Section 5.2); but we would like to mention just one way in which the current account might be generalized.

It is a common objection to deontic logics of this standard kind that they 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. In fact, however, our semantic framework can

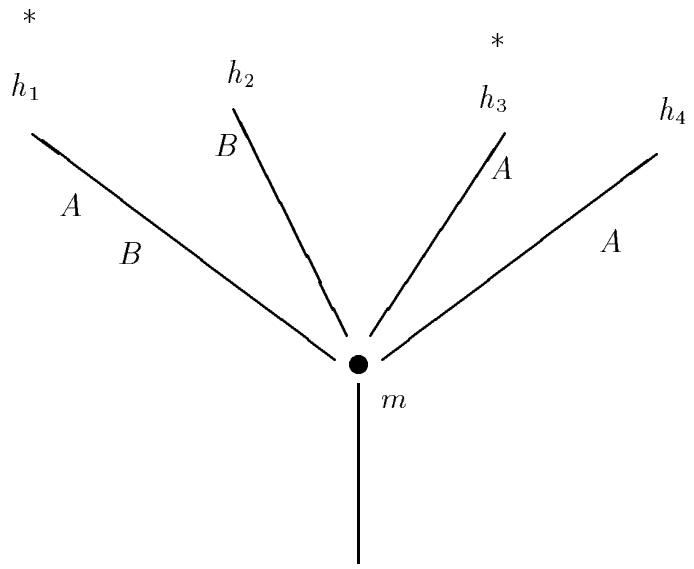

Figure 9: $\mathit{Ought}\,(m) = \{h_1, h_3\}$.

be generalized to accommodate a broader range of normative theories. Let us replace 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 a set of values. The value assigned to a history h at m represents the worth or desirability of that history at m; and we assume that the space of values is partially ordered by $\leq$, so that $\mathit{Value}_m(h) \leq \mathit{Value}_m(h')$ means that h' has a value at m at least as great as that of h. In these new frames, the above evaluation rule for ought statements can now be replaced with the following:

- $\mathcal{M}, m/h \models \bigcirc A$ iff there is a 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 $\mathit{Value}_m(h') \leq \mathit{Value}_m(h'')$.

This revised evaluation rule is, in fact, a conservative generalization of the previous version: if we take $\mathit{Value}_m(h) = 1$ just in case $h \in \mathit{Ought}\,(m)$, and $\mathit{Value}_m(h) = 0$ otherwise, and we suppose that the values are ordered so that $0 < 1$, then the two definitions will generate the same ought statements. However, the revised rule can apply also to normative theories that allow for more than two values, and in which the ordering among values is more complicated. For example, the rule can apply to utilitarian theories, which would associate with each history passing through a moment, as its value, a real number representing the

utility (or expected utility) of that history at that moment.

In the utilitarian case, although there are a number of values, they still stand, of course, in a linear ordering; but the revised evaluation rule can also accommodate more radical departures from standard deontic logic in which even the assumption of a linear ordering among values is dropped, and in which statements of the form $\bigcirc A \wedge \bigcirc \neg A$ are satisfiable. As an example, van Fraassen [41] describes a nonstandard deontic logic in which oughts are founded on background sets of imperatives; an ought statement is thought of as true if it is a necessary condition for satisfying some maximal set of imperatives. We could incorporate this idea into the present context 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 "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; and let us take the set of values as ordered by subset inclusion. It then turns out, as noted in [26], that the oughts generated by the revised evaluation rule coincide with those supported by van Fraassen's own definition.

In spite of these possibilities for generalization, we will work in the remainder of this paper with the simpler deontic framework that classifies histories only as ideal or non-ideal.

5.2 Oughts and the deliberative stit

The 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 along all of the ideal histories, it will not snow; 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 ability (personal can-do) earlier through a combination of historical possibility (impersonal can) and the deliberative stit (personal to-do), we might hope to arrive at a theory of what an agent ought to do in the same way: by combining the deliberative stit with our impersonal account of what ought to be.

This general strategy was favored by a number of German writers toward the beginning of the century, notably Meinong and Nicolai Hartmann; and the strategy has been explicitly

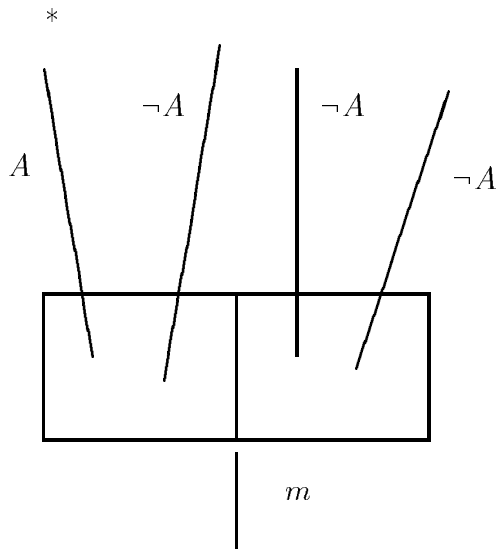

Figure 10: $\bigcirc A$ without $\bigcirc[\alpha\ \mathit{dstit}\colon A]$.

endorsed by at least one contemporary: Chisholm suggests in [15, 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."[19] 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 found already in the writings of St. Anselm. But we can follow the same general strategy, relying instead on the historical ought and the deliberative stit. The result is a proposal that the formula

$$\bigcirc[\alpha\ \mathit{dstit}\colon A]$$

can be taken to express the claim that α ought to see to it that A, or that α is obligated to see to it that A.

In the present context of branching time, 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 ideal histories belonging to $\mathit{Ought}\,(m)$ filter through the Choice^m_α partition. Consider, for example, the situation depicted in Figure 10. Here, $\bigcirc A$ is settled true at m. However, $\Diamond[\alpha\ \mathit{dstit}\colon A]$ 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, we

[19] Chisholm's paper contains a reference to Hartmann's work; a recent discussion of Meinong's proposal can be found in García [21].

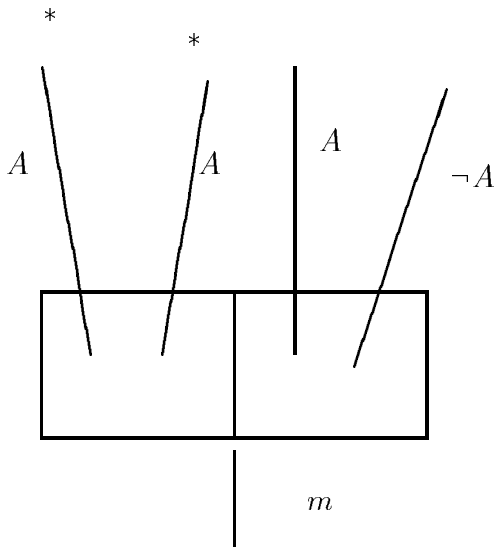

Figure 11: $\bigcirc[\alpha\ \mathit{dstit}\colon A]$.

know that

$$\bigcirc[\alpha\ \mathit{dstit}\colon A] \supset \Diamond[\alpha\ \mathit{dstit}\colon A],$$

or that obligation implies ability—α can be obliged to see to it that A only if he is able to do so. Because α is unable at m to see to it that A, we can thus conclude that $\bigcirc[\alpha\ \mathit{dstit}\colon A]$ is settled false there as well. By contrast, Figure 11 depicts a situation in which $\bigcirc[\alpha\ \mathit{dstit}\colon A]$ is settled true: $[\alpha\ \mathit{dstit}\colon A]$ holds at m/h for each history h belonging to $\mathit{Ought}\,(m)$.

Let us now consider some possible theses concerning what, on the present analysis, an agent ought to do.

Since $\bigcirc$ is a a normal modal operator, we know that any statement of the form $(A_1 \wedge \ldots \wedge A_n) \supset B$ implies a corresponding statement of the form $(\bigcirc A_1 \wedge \ldots \wedge \bigcirc A_n) \supset \bigcirc B$; and so it is a straightforward matter to derive principles concerning the logical properties of an agent's obligations from principles such as those established in Section 3.1 concerning the logical properties of the deliberative stit. We can conclude from the deliberative stit validity of the theses C and SMP, for example, that what an agent ought to do is closed under conjunction and modus ponens: the statements

$$\bigcirc[\alpha\ \mathit{dstit}\colon A] \wedge \bigcirc[\alpha\ \mathit{dstit}\colon B] \supset \bigcirc[\alpha\ \mathit{dstit}\colon A \wedge B],$$
$$\bigcirc[\alpha\ \mathit{dstit}\colon A] \wedge \bigcirc[\alpha\ \mathit{dstit}\colon A \supset B] \supset \bigcirc[\alpha\ \mathit{dstit}\colon B]$$

are both valid.

On the other hand, we lose certain validities of standard deontic logic when the notion of what ought to be the case is replaced by the notion of what some agent ought to do. For example, standard deontic logic yields validities such as

$$\bigcirc A \supset \bigcirc(A \vee B),$$
$$\bigcirc(A \wedge B) \supset \bigcirc A.$$

Opinions are split as to the intuitive desirability of formulas like these, which register the closure of the ordinary $\bigcirc$ operator under logical consequence. The first of these statements, for instance, is often taken as a schematic expression of Ross's paradox: if an agent ought to mail a letter, then it follows that he ought either to mail the letter or burn it.[20] To some writers, this has seemed like an awkward implication to endorse; but to others it has seemed benign, and even natural: if it is a necessary condition for achieving an ideal state that the letter should be mailed, then it is a necessary condition for achieving such a state that the letter should be either mailed or burned.

Without examining the matter in great detail, we wish to suggest that there may be a sense in which the present proposal allows us to endorse both sides of this issue. Since the validities of standard deontic logic carry over into the present context, we must accept the two formulas listed above; and so we can agree with those who feel that these statements express natural conditions on what ought to be. However, the analogous formulas

$$\bigcirc[\alpha \; \mathit{dstit}\colon A] \supset \bigcirc[\alpha \; \mathit{dstit}\colon A \vee B],$$
$$\bigcirc[\alpha \; \mathit{dstit}\colon A \wedge B] \supset \bigcirc[\alpha \; \mathit{dstit}\colon A]$$

are both invalid; and so we can agree also with those who find an awkwardness in Ross's paradox. Because the deliberative stit itself fails to to satisfy closure under consequence, we can accept the consequential closure of what ought to be while still denying the consequential closure of what an agent ought to do. We can agree that if a letter ought to be mailed, then

[20] General discussions of Ross's paradox, with references to the literature, can be found in Føllesdal and Hilpinen [19, pp. 21–23], and also in in Åqvist [1, pp. 634–638]; the issue is discussed from the perspective of stit semantics in Perloff [33]. Some problems with the second formula listed above are noted in von Wright [46, pp. 7–8].

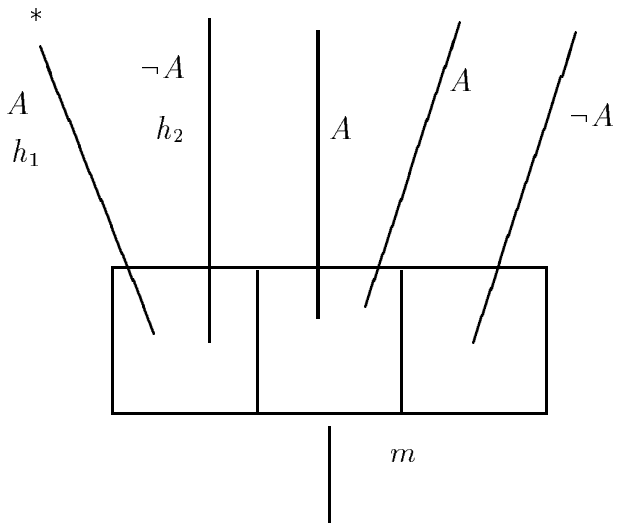

Figure 12: $\bigcirc A$ and $\Diamond[\alpha\ \mathit{dstit}\colon A]$ without $\bigcirc[\alpha\ \mathit{dstit}\colon A]$.

it ought to be either mailed or burned; but we are not forced to conclude that if an agent ought to mail a letter, then he ought either either to mail it or burn it.

One advantage of the present proposal—which abandons the Restricted Complement Thesis, and instead analyzes what an agent ought to do through a combination of the deliberative stit and a generally applicable deontic operator—is that the resulting theory then provides a framework for studying the interactions between what an agent ought to do and what ought to be. It is easy to see, for example, that the statement

$$\bigcirc[\alpha\ \mathit{dstit}\colon A] \supset \bigcirc A$$

is valid in the present context: if α ought see to it that A, then it ought to be that A. This seems a cheering result to some (including at least one of us), since it means that the agent is never obliged to waste his time bringing about a state of affairs that, in itself, need not hold.[21]

[21] This principle has been challenged, however, by Fred Feldman [18, p. 193], who presents a semantic theory that allows for a counterexample. Although Feldman's semantic theory shares a number of features with that developed here, we have been unable to interpret the counterexample he describes within our present framework; we suspect that it relies on a deterministic assumption running against the indeterministic tense logic that forms the background of our investigation.

We noted earlier the failure of the converse implication,

$$\bigcirc A \supset \bigcirc[\alpha \; \mathit{dstit}\colon A];$$

even if it ought to be that A, the agent may be under no obligation to bring about such a state of affairs. But let us look a bit more closely at the countermodel set out earlier, in Figure 10, to demonstrate the invalidity of this formula. The example involved a situation in which α did not even have the ability to see to it that A—the statement $\Diamond[\alpha \; \mathit{dstit}\colon A]$ failed—and so, since obligation implies ability, we were able to conclude at once that $\bigcirc[\alpha \; \mathit{dstit}\colon A]$ should fail as well. It might appear that this kind of countermodel actually depends on the agent's lack of ability, so that a weaker statement of the form

$$\bigcirc A \wedge \Diamond[\alpha \; \mathit{dstit}\colon A] \supset \bigcirc[\alpha \; \mathit{dstit}\colon A]$$

might hold: if it ought to be that A, and the agent is able to bring it about that A, then he ought to do so. It turns out, however, that even this weaker implication is invalid, as we can see from Figure 12. Here, both $\bigcirc A$ and $\Diamond[\alpha \; \mathit{dstit}\colon A]$ are settled true at m: it ought to be that A, and the agent is able to bring it about. But $\bigcirc[\alpha \; \mathit{dstit}\colon A]$ is settled false, and indeed we have $\bigcirc[\alpha \; \mathit{dstit}\colon \neg[\alpha \; \mathit{dstit}\colon A]]$ settled true: the agent has no obligation to bring it about that A, and in fact is obliged to refrain from doing so.

Although formally transparent, the situation depicted in Figure 12 is complicated enough conceptually that it is worth fleshing out the abstract model with a story. So suppose that the agent, Lucinda, wishes to buy a horse, but that she has only \$10,000 to spend and the horse she wants is selling for \$15,000. We imagine that Lucinda offers \$10,000 for the horse at the moment m, choosing the leftmost cell in the choice partition, and that the matter is then out of her hands; it is up to the owner of the horse to decide whether to accept the offer. The history h_1 represent a scenario in which the owner accepts Lucinda's offer, h_2 a scenario in which the offer is rejected, and A the statement that Lucinda will become less wealthy by \$10,000. The unique ideal history, we will suppose, is h_1, in which the offer is accepted, and, as a consequence, Lucinda buys the horse and becomes less wealthy by \$10,000. Since Lucinda is out \$10,000 in the ideal history, we must conclude that it ought

to be that she is out \$10,000. Of course, Lucinda is also able to see to it that she is out \$10,000, as depicted in the middle cell of the choice partition—perhaps by throwing the cash down a storm drain. But we should not conclude that Lucinda ought to see to it that she is out \$10,000; and in fact, in the ideal history, she refrains from doing so.

As a final example of the relation between what ought to be and what, on the present analysis, an agent ought to do, we just mention the pleasing validity of

$$\bigcirc(\bigcirc[\alpha \; \mathit{dstit}\colon A] \supset [\alpha \; \mathit{dstit}\colon A]),$$

which says, of course, that it ought to be that an agent does what he ought to do. The formula does not depend on any features of the deliberative stit at all, but is simply an instance of the validity $\bigcirc(\bigcirc B \supset B)$.

5.3 Ought to do

So far we have been exploring the consequences of the suggestion, found in Meinong and Chisholm, that the content of a statement of the form "α ought to see to it that A" can be captured through a statement of the form "It ought to be that α sees to it that A." We now turn from exploring the consequences of this suggestion to evaluating the suggestion itself. Is it reasonable to suppose that, at a fundamental level, it is situations that are classified as good or bad, and that personal obligations are then derived from these?

The general idea of deriving personal from impersonal oughts in this way has been subject to several logical objections; but in fact, when it is deployed in the current context of stit semantics, much of this criticism can be deflected.

One line of objection—due to Peter Geach [22], who traces it to St. Anselm—is that the proposal appears to render judgments about what agents ought to do inappropriately insensitive to transformations in the complement of the ought. The argument proceeds as follows. Suppose, for example, that Fred ought to dance with Ginger, that Fred is obligated to do so. According to the suggested analysis, this should be taken to mean that it ought to be that Fred dances with Ginger. Now as it happens, the relation of dancing with is symmetric. In any possible world in which Fred dances with Ginger, Ginger dances also with

Fred, and vice versa; and so it seems that the two statements "Fred dances with Ginger" and "Ginger dances with Fred" are necessarily equivalent. It follows from standard deontic logic that if the statements A and B are necessarily equivalent, then the statement $\bigcirc A$ is likewise equivalent to $\bigcirc B$. We can thus conclude, since it ought to be that Fred dances with Ginger, that it ought to be also that Ginger dances with Fred; and then according to the suggested analysis, again, this would lead us to conclude that Ginger ought to dance with Fred. But of course, this conclusion is incorrect: it could easily happen that, because of the customs governing some social occasion, Fred is obligated to dance with Ginger even though Ginger is not obligated to dance with Fred.[22]

In the context of stit semantics, which provides a framework for reasoning explicitly about agency, this kind of objection can be met simply by noting that the argument fails to consider whose agency is involved in the complement of the ought. Let A represent the statement that Fred and Ginger dance together; and let α and β represent Fred and Ginger respectively. In the present context, the statement that Fred ought to dance with Ginger would be analyzed through the formula $\bigcirc[\alpha\ \mathit{dstit}\colon A]$, in which Fred's agency is explicit: it ought to be that Fred sees to it that he and Ginger dance together. Of course, the formulas $[\alpha\ \mathit{dstit}\colon A]$ and $[\beta\ \mathit{dstit}\colon A]$—that Fred sees to it that he and Ginger dance, and that Ginger sees to it that she and Fred dance—are not equivalent. And so there is no reason to draw the conclusion $\bigcirc[\beta\ \mathit{dstit}\colon A]$, that Ginger ought to see to it that she and Fred dance; or on the present analysis, that Ginger ought to dance with Fred.

The situation can be illustrated as in Figure 13. Here, α's choices at m are represented by the vertical partition of $H_{(m)}$, and β's by the horizontal partition; A is the statement that the two agents will dance together; and h_1 is the unique ideal history at m, the unique history belonging to $\mathit{Ought}\,(m)$. Evidently, $\bigcirc[\alpha\ \mathit{dstit}\colon A]$ is settled true at m—α ought to

[22]This is not Geach's own example. Geach himself develops the objection by arguing that the statement "Tom ought to be beaten up by John" might be true even though "John ought to beat up Tom" is false, and even though "Tom is beaten up by John" seems to be equivalent to "John beats up Tom." We change the example not just for the sake of delicacy, but also because we agree with García [21] that Geach's own example introduces considerations concerning the logic of just desert that are irrelevant to the matter at hand.

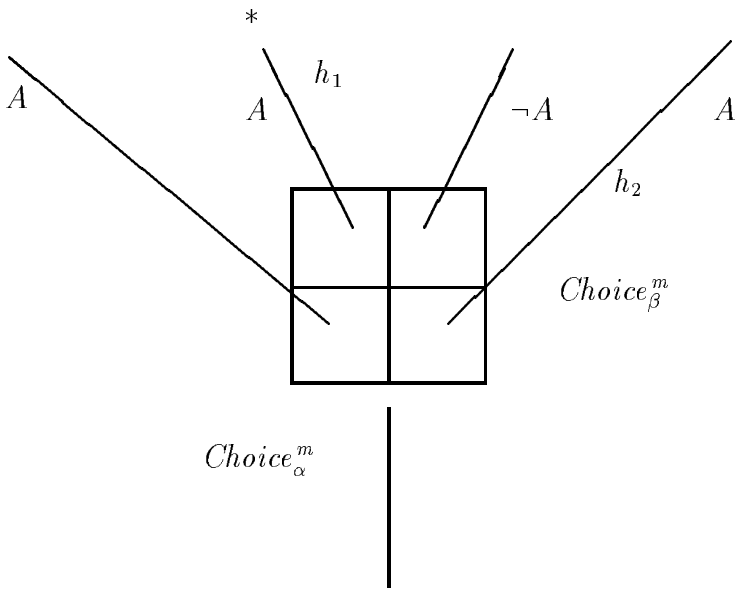

Figure 13: $\bigcirc[\alpha\ dstit\colon A]$ without $\bigcirc[\beta\ dstit\colon A]$.

see to it that the agents dance together—since $[\alpha\ dstit\colon A]$ is true at the index m/h_1. Of course, the agent β also has the ability to see to it that they dance together; $\Diamond[\beta\ dstit\colon A]$ is settled true at m, since $[\beta\ dstit\colon A]$ is true at m/h_2. But $\bigcirc[\beta\ dstit\colon A]$ is settled false. Even though A represents something that ought to happen, and that α ought to see to, and that β has the ability to see to, it is not something that β ought to see to. In fact, it is something that β ought to refrain from seeing to; the formula $\bigcirc[\beta\ dstit\colon \neg[\beta\ dstit\colon A]]$ is settled true at m, since $[\beta\ dstit\colon \neg[\beta\ dstit\colon A]]$ holds at m/h_1.

Another, related objection to the idea of reducing personal to impersonal oughts in the way suggested here can be found in Gilbert Harman [23, Appendix B], who bases his objection on a distinction, pointed out originally by I. L. Humberstone [27], between two kinds of ought statements—what he calls "situational" and "agent-implicating" oughts.

To illustrate the distinction, let us first imagine a case in which Albert has competed in a gymnastics event. Suppose Albert's performance is clearly superior, but the judge is known to be biased, and it is likely that he will award the medal to someone else. If one then said, "Albert ought to win the medal," this is the kind of statement that Humberstone would classify as a situational ought. It reflects a judgment about the situation, not about Albert,

and can legimately be paraphrased as "It ought to be that Albert wins the medal." There is no implication that Albert will be at fault if he fails to win the medal, or that winning the medal is now within his power. By contrast, suppose Albert has not kept up with his training schedule. One might then say, "Albert ought to practice harder," and this would be the kind of ought statement that Humberstone classifies as agent-implicating. It implies that Albert is able to practice harder, and places the blame on him if he fails to do so.

Now Harman's objection to the present suggestion for analyzing personal oughts in terms of impersonal oughts is simply that it obscures the distinction between the two kinds of ought statements that Humberstone has identified. According to the suggestion, the statement "Albert ought to practice harder" is itself to be analyzed as "It ought to be that Albert practices harder." It is thus similar in form to "It ought to be that Albert wins the medal," and so it is hard to see why one of these ought statements should be classified as agent-implicating and the other as situational.

Again, however, the objection can by met without abandoning the present suggestion for analyzing personal oughts, by focusing, not on the relation between the ought operator and its complement, but instead on the treatment of agency within its complement. In those ought statements that Humberstone regards as agent-implicating, the complement should be represented as a stit sentence. For example, the statement that Albert ought to practice harder should be represented through the formula $\bigcirc[\alpha\ dstit\colon A]$, where α is Albert, and A is the statement that he practices harder. As we have seen, this implies $\Diamond[\alpha\ dstit\colon A]$, that Albert is able to practice harder. On the other hand, in a situational ought, the complement will not be agentive. The idea that Albert ought to win the medal might be represented as $\bigcirc B$, where B is the statement simply that he wins the medal, that it is awarded to him, and not the statement that he sees to it that he wins the medal. Of course, $\bigcirc B$ alone does not imply $\Diamond[\alpha\ dstit\colon B]$, that Albert is able to see to it that he wins the medal.

In the current context of stit semantics, then, Meinong's idea of deriving personal from impersonal oughts—analyzing a statement of the form "α ought to see to it that A" through a statement of the form "It ought to be that α sees to it that A"—is able to withstand at least the kind of logical objections deployed by Geach and Harman.

And in fact, this analysis does seem to capture the notion of personal obligation at work in certain ethical theories—particularly, pure consequentialist theories, such as act utilitarianism. On these theories, the only real source available for supporting personal obligations is an independent notion of the value assigned to particular states of affairs. What one ought to do in any given situation, according to theories of this kind, is simply bring about a maximally valuable state of affairs from among those achievable in that situation; and if performing a certain action is a necessary condition for bringing about such a state of affairs, then the agent is obligated to perform that action. Evidently, this consequentialist analysis of personal obligation corresponds closely to Meinong's idea: both identify what an agent is obligated to do in a given situation with what it would be best, or most valuable, for the agent to do.

It is exactly this aspect of consequentialist theories, however, that exposes them to the standard kind of counterarguments, and that leads so many people to find these theories so implausible. For example, once an agent's income reaches a certain level, and if the agent is unemcumbered by extraordinary expenses, one might think it best for the agent to donate some percentage of his income to charity; this may be something that is necessary for maximizing overall utility, and that happens in all the ideal worlds. Still, most people would resist the conclusion that the agent has an obligation to contribute to charity. Even if it is best for the agent to contribute to charity—even if this is a necessary condition for achieving a maximally valuable state of affairs—it does not seem to follow that this is something the agent is obligated to do.

Most of us are not consequentialists at heart: our ethical intuitions allow for a distinction between what it would be best for an agent to do and what the agent is obligated to do. In order to accommodate this distinction in the current framework, it would be necessary to abandon the idea of deriving personal from impersonal oughts, and attempt instead to provide an independent analysis of what an agent ought to do.

5.4 Indexed ought sets

We do not attempt in this paper to offer any such independent analysis of the concept of personal obligation, what an agent ought to do. But we do wish in this final section to explore one initially attractive approach to the concept in the context of the deliberative stit.

One problem with the idea of reducing personal to impersonal oughts, grounding the notion of what an agent ought to do in an agent-independent notion of what ought to be, is that the latter notion relies at any given moment on a single set of ideal possibilities, a single ought set, applicable to everyone. Yet it seems plain that our individual obligations vary—depending, for example, upon our individual roles in society, or the commitments we have individually undertaken. Of course, even the theory developed in Section 5.2, which does ground personal oughts in impersonal oughts, allows different agents at the same moment to face different personal obligations: as we saw in the model depicted in Figure 13, it might be, even according to this theory, that α is obligated to see to it that A while β is not. Still, this kind of variance among the personal obligations of different agents results only from the different ways in which the histories classified as ideal at a moment might filter through the different choice partitions of these agents; and it seems that there should be more to it than that. It seems that the set of histories classified as ideal—those in which things turn out as they ought to—might itself vary from one agent to another.

It is a straightforward matter to incorporate this idea into the current framework. The frames described in Section 5.1 could be modified so as to contain, rather than a unary function mapping each moment m into a subset $\mathit{Ought}\,(m)$ of $H_{(m)}$, instead a binary function mapping each agent α and moment m into a nonempty subset $\mathit{Ought}_{\alpha}\,(m)$ of $H_{(m)}$, interpreted as the set of histories at m that are ideal from the point of view of the agent α. The language could then be modified to include, rather than a single deontic operator, a set of indexed operators, one for each agent ($\bigcirc_{\alpha}$, $\bigcirc_{\beta}$, ...), with the following evaluation rule:

- $\mathcal{M}, m/h \models \bigcirc_{\alpha} A$ iff $\mathcal{M}, m/h' \models A$ for each $h' \in \mathit{Ought}_{\alpha}\,(m)$.

The logic of each of these individual indexed ought operator would then coincide with the

logic of our previous unindexed ought; but because there are no constraints on the relations among the ought sets of different agents, the logic of formulas with mixed indices would have a nonstandard flavor. For example, $Ought_{\alpha}(m)$ and $Ought_{\beta}(m)$ could easily be arranged so that $\bigcirc_{\alpha}A \wedge \bigcirc_{\beta}\neg A$ holds at m.

The idea of indexing deontic operators in this way is not new, of course; it is found, for example, in Thomason:

> Deontic logicians have had a habit of speaking as if there were a single $\bigcirc$, and the formalization of 'John ought to apologize to Jane will be $\bigcirc Pa$ (where Pu formalizes '...apologizes to Jane') and that of 'Bill ought to apologize to Jane' will be $\bigcirc Pb$. Semantically this would mean one and the same ought set, serving for everyone. This is wrong; everyone should have his own ought set, and the formalization of our two sentences should be $\bigcirc_{a}Pa$ and $\bigcirc_{b}Pb$ [38, p. 183].

But in itself, simply indexing the ought sets to the different agents, and then supplementing the language with indexed deontic operators, is not really sufficient to capture the notion of personal obligation, what an agent ought to do. There is no requirement that these indexed deontic operators should apply only to formulas representing actions, or that the index of the deontic operator should coincide with the agent of the action.

Let us, however, impose both of these requirements. The most natural way to do this is to proceed in two steps. First, we return to the Restricted Complement Thesis, set aside earlier, in Section 5.1. According to this thesis, deontic operators must take stit formulas as their complements; the thesis thus requires an ought operator to occur only in a context of the form $\bigcirc[\alpha\ dstit{:}\ A]$, and so approximates the requirement that such an operator should apply only to sentences representing actions. Second, we imagine that, in such a context, the index of the deontic operator is supplied implicitly by the agent of the stit formula to which it is affixed, so that, of course, index and agent would have to coincide.

The effect of these restrictions is that a sentence $\bigcirc[\alpha\ dstit{:}\ A]$ can be regarded as formed from the matrix A, not through successive applications first of the deliberative stit and then of a generally applicable deontic operator, but instead, through an application of the

single, fused connective $\bigcirc[\alpha\ dstit\colon \ldots]$, meaning "$\alpha$ is obligated to see to it that ...". The evaluation rule for this fused connective would then be:

- $\mathcal{M}, m/h \models \bigcirc[\alpha\ dstit\colon A]$ iff for each $h' \in Ought_{\alpha}(m)$ we have (1) $\mathcal{M}, m/h'' \models A$ for each $h'' \in Choice_{\alpha}^{m}(h')$, and (2) $\mathcal{M}, m/h''' \not\models A$ for some $h''' \in H_{(m)}$.

Of course, this rule gives the statement $\bigcirc[\alpha\ dstit\colon A]$ truth conditions equivalent to those of the statement formed by affixing an indexed ought $\bigcirc_{\alpha}$ to the formula $[\alpha\ dstit\colon A]$. However, in keeping with the Restricted Complement Thesis, it does so without assigning any independent meaning to the ought operator, and without allowing us to evaluate sentences of the form $\bigcirc_{\alpha}B$ in which B is not of the form $[\alpha\ dstit\colon A]$.

This strategy for analyzing personal obligation—adhering to the Restricted Complement Thesis, and then appealing to indexed ought sets in evaluating obligation statements—was first explored by Paul Bartha in [2], and then developed further in Belnap and Bartha [5]. In fact, the fused connective just defined, with the evaluation rule provided, is equivalent (under certain natural conditions) to the connective set out by Bartha for expressing personal obligation.

We now have before us, therefore, two approaches to the task of analyzing the notion of personal obligation, what an agent ought to do. The first is the approach based on Meinong's idea of regarding a statement of the form "α is obligated to see to it that A" as equivalent ot a statement of the form "It ought to be that α sees to it that A." As developed earlier, in Section 5.2, this approach treats statements of the form $\bigcirc[\alpha\ dstit\colon A]$ as resulting from a combination of a generally applicable, agent-independent ought operator $\bigcirc$ with the stit formula $[\alpha\ dstit\colon A]$. The second approach, developed in the present section, combines the Restricted Complement Thesis with indexed deontic operators: the ought operator can occur only in the context $\bigcirc[\alpha\ dstit\colon A]$, and in such a context it is interpreted semantically against the ought set indexed to the agent α.

We can now compare these two approaches. Of course, the first approach leads to a language that is syntactically more expressive, since it is not governed by the Restricted Complement Thesis, and allows the ought operator to apply to arbitrary formulas. But

let us limit our attention to the common fragment of the two languages, that in which the ought operator applies only to stit formulas. We can then compare validities. Once we have limited our consideration to their common linguistic fragment, the only difference between the two approaches is that, according to the second, the ought operator is interpreted as indexed, while it is unindexed according to the first. As we have seen, the indexing of generally applicable deontic operators leads to the satisfiability of certain formulas—such as $\bigcirc_\alpha A \wedge \bigcirc_\beta \neg A$—whose unindexed analogs are not satisfiable. One might expect (and we expected) to find a similar sort of difference between the two approaches under consideration here. Surprisingly, however, this is not what happens. Once one restricts the indexed oughts to take only stit sentences as their complements, it then turns out—as we show in the Appendix to this paper—that exactly the same formulas are validated according to each of the two approaches, whether the oughts are indexed or not.

There may be in sense, then, in which the approach developed in the present section results in a construction that is closer to our notion of personal obligation than the approach developed earlier, in Section 5.2. But the two approaches cannot be distinguished by looking only at the validities supported in their common language; validities alone do not tell us to what extent, if at all, the present approach takes us beyond Meinong's analysis, which reduces personal to impersonal oughts. For this reason, the representation of personal obligation, what an agent ought to do, must remain a matter for further exploration.[23]

A A result about indexed oughts

This appendix establishes the technical fact referred to at the end of Section 5.4. If we adopt the Restricted Complement Thesis—limiting the $\bigcirc$ operator so that it occurs only in contexts of the form $\bigcirc[\alpha\ \mathit{dstit}\colon A]$—then at least in terms of validities, it makes no difference whether or not the ought sets used for interpreting this operator are indexed to

[23] A different approach to explicating the notion of personal obligation within the general framework of stit semantics is developed in [25].

agents: exactly the same formulas are validated either way.[24]

We begin with some definitions. Let an *ought frame* be a structure of the form

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

as described in Section 5.1. Let an *indexed ought frame* be a structure like an ought frame except that, as described in Section 5.4, the unary function mapping each moment m into a nonempty subset $Ought(m)$ of $H_{(m)}$ is replaced by a binary function mapping each agent α and moment m into a nonempty subset $Ought_\alpha(m)$ of $H_{(m)}$. *Ought models* and *indexed ought models* result from interpreting the background language against ought frames and indexed ought frames, respectively.

For the purpose of comparison, it will be useful to distinguish the turnstiles representing the satisfaction relations associated with these different kinds of models: we let an ordinary $\models$ represent the satisfaction relation associated with ought models, and an indexed $\models_i$ represent the satisfaction relation associated with indexed ought models. These two satisfaction relations agree in interpreting all but the deontic connectives as explained throughout this paper. The $\models$ relation interprets the deontic operator according to the ordinary evaluation rule

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

from Section 5.1. The $\models_i$ relation appeals officially to the Section 5.4 rule

- $\mathcal{M}, m/h \models_i \bigcirc[\alpha\ dstit\text{: } A]$ iff for each $h' \in Ought_\alpha(m)$ we have (1) $\mathcal{M}, m/h'' \models_i A$ for each $h'' \in Choice^m_\alpha(h')$, and (2) $\mathcal{M}, m/h''' \not\models_i A$ for some $h''' \in H_{(m)}$.

However, the reader can easily see that this official statement of the rule is equivalent to the simpler formulation

- $\mathcal{M}, m/h \models_i \bigcirc[\alpha\ dstit\text{: } A]$ iff for each $h' \in Ought_\alpha(m)$ we have $\mathcal{M}, m/h' \models_i [\alpha\ dstit\text{: } A]$.

[24]The proof contained in this appendix evolved after discussions with Paul Bartha, who suggested several of the key ideas to us.

We will rely on this simpler formulation in this appendix.

We define a formula A as *satisfiable in ought models* if there is an ought model $\mathcal{M}$ such that $\mathcal{M}, m/h \models A$, and *satisfiable in indexed ought models* if there is an indexed ought model $\mathcal{M}$ such that $\mathcal{M}, m/h \models_i A$. A formula A is *valid* in ordinary or indexed ought models, respectively, if $\neg A$ is not satisfiable.

In what follows, we will consider only sentences formed in accord with the Restricted Complement Thesis—that is, sentences in which all the other connectives discussed in this paper (the truth-functional connectives, P, F, $\Box$, and *dstit*) can occur freely, but in which $\bigcirc$ occurs only in contexts of the form $\bigcirc[\alpha\ \mathit{dstit}\colon A]$. The central result of this appendix, then, is:

Theorem 1 *A sentence A is satisfiable in ought models iff A is satisfiable in indexed ought models.*

And of course, from this it follows at once that ought models and indexed ought models support the same validities.

The proof of this Theorem relies on several preliminary definitions and lemmas. We first define a mapping $[\ldots]_i$ from ought models to indexed ought models, in the following way. Where $\mathcal{M}$ is an ought model, we let $[\mathcal{M}]_i$ be the indexed ought model otherwise like $\mathcal{M}$ but in which, for each agent α, we have $h \in \mathit{Ought}_\alpha(m)$ in $[\mathcal{M}]_i$ just in case $h \in \mathit{Ought}(m)$ in $\mathcal{M}$. That is, in moving from $\mathcal{M}$ to $[\mathcal{M}]_i$, at each moment, the single impersonal ought set from $\mathcal{M}$ is assigned to each agent as that agent's personal ought set. We can now establish:

Lemma 1 *Let $\mathcal{M}$ be an ought model. Then for each m/h pair and for each sentence A, we have $\mathcal{M}, m/h \models A$ iff $[\mathcal{M}]_i, m/h \models_i A$.*

Proof. By induction on the complexity of the formula A. Since $\mathcal{M}$ and $[\mathcal{M}]_i$ agree everywhere but in their treatment of deontic operators, the only interesting case is that in which A has the form $\bigcirc[\alpha\ \mathit{dstit}\colon B]$. So suppose, first, that $\mathcal{M}, m/h \models \bigcirc[\alpha\ \mathit{dstit}\colon B]$. Then $\mathcal{M}, m/h' \models [\alpha\ \mathit{dstit}\colon B]$ for each $h' \in \mathit{Ought}(m)$. From the definition of $[\mathcal{M}]_i$, we can conclude that $\mathcal{M}, m/h' \models [\alpha\ \mathit{dstit}\colon B]$ for each $h' \in \mathit{Ought}_\alpha(m)$. By inductive hypothesis, we can then conclude that $[\mathcal{M}]_i, m/h' \models_i [\alpha\ \mathit{dstit}\colon B]$ for each $h' \in \mathit{Ought}_\alpha(m)$, so

that $[\mathcal{M}]_i, m/h \models_i \bigcirc[\alpha\ dstit{:}\ B]$. The argument that $[\mathcal{M}]_i, m/h \models_i \bigcirc[\alpha\ dstit{:}\ B]$ only if $\mathcal{M}, m/h \models \bigcirc[\alpha\ dstit{:}\ B]$ is similar. ∎

From Lemma 1, we can see that Theorem 1 holds from left to right: if $\mathcal{M}, m/h \models A$ for the ought model $\mathcal{M}$, then $[\mathcal{M}]_i, m/h \models_i A$ for the indexed ought model $[\mathcal{M}]_i$.

In order to establish the other direction of the Theorem, we define a mapping from indexed ought models into ordinary ought models that can be shown to satisfy the same sentences at every point of evaluation. Supposing, then, that $\mathcal{M}$ is an indexed ought model, we move through the following three steps:

(1) Beginning with the *Ought* function from $\mathcal{M}$, define a new function $Ought'$ so that, for each agent α and moment m, we have the history $h \in Ought'_\alpha(m)$ just in case there is a history h' such that $h' \in Ought_\alpha(m)$ and $h \in Choice^m_\alpha(h')$. Informally, this means that $Ought'_\alpha(m)$ expands $Ought_\alpha(m)$ so as to contain every history from each choice cell K belonging to the partition $Choice^m_\alpha$ whenever $Ought_\alpha(m)$ contains any history belonging to K. It is useful to note that $Ought'_\alpha(m)$ can thus be characterized as a union of certain cells belonging to $Choice^m_\alpha$:

$$Ought'_\alpha(m) = \bigcup\{K : K \in Choice^m_\alpha \text{ and } K \cap Ought_\alpha \neq \emptyset\}.$$

(2) For each moment m, define the impersonal ought set $Ought^*(m)$ so that $h \in Ought^*(m)$ just in case $h \in Ought'_\alpha(m)$ for each agent α. Since $Ought^*(m)$ is thus simply the intersection of the various $Ought'_\alpha(m)$ sets for each agent α, we can see from the above characterization of these individual $Ought'_\alpha$ sets that

$$Ought^*(m) = \bigcap_{\alpha \in Agent} \bigcup\{K : K \in Choice^m_\alpha \text{ and } K \cap Ought_\alpha \neq \emptyset\}.$$

(3) Define an ordinary ought model $\mathcal{M}^*$ by replacing the binary, personal ought function *Ought* from the indexed ought model $\mathcal{M}$ with the unary, impersonal ought function $Ought^*$.

In order to verify that each formula satisfiable in indexed ought models is satisfiable in ordinary ought models—completing the proof of Theorem 1—we now want to show that

the ordinary ought model $\mathcal{M}^*$ supports the same statements as our original indexed ought model $\mathcal{M}$ at any point of evaluation:

$\mathcal{M}, m/h \models_i A$ iff $\mathcal{M}^*, m/h \models A$.

As a means of accomplishing this task, it is convenient, as a fourth step, to introduce yet another indexed model:

(4) From the ordinary ought model $\mathcal{M}^*$ introduced above, define the indexed ought model $[\mathcal{M}^*]_i$ as specified by the $[\ldots]_i$ function, by taking $h \in \mathit{Ought}^*_\alpha(m)$ for each agent α just in case $h \in \mathit{Ought}^*(m)$. Since, for each agent α, the set $\mathit{Ought}^*_\alpha(m)$ simply coincides with $\mathit{Ought}^*(m)$, we can thus conclude from the characterization of $\mathit{Ought}^*(m)$ in step (2) that

$$\mathit{Ought}^*_\alpha(m) = \bigcap_{\alpha \in Agent} \bigcup \{K : K \in \mathit{Choice}^m_\alpha \text{ and } K \cap \mathit{Ought}_\alpha \neq \emptyset\}.$$

From Lemma 1, we know that the indexed model $[\mathcal{M}^*]_i$ must agree with $\mathcal{M}^*$ at every point of evaluation:

$[\mathcal{M}^*]_i, m/h \models_i A$ iff $\mathcal{M}^*, m/h \models A$.

And so it suffices to show that, at every point of evaluation, the two indexed models $[\mathcal{M}^*]_i$ and $\mathcal{M}$ must agree:

$\mathcal{M}, m/h \models_i A$ iff $[\mathcal{M}^*]_i, m/h \models_i A$.

This latter fact is verified in Lemma 4, below; but the proof of this Lemma depends upon two others, one whose proof is omitted as obvious, and one whose proof we only sketch.

Lemma 2 *For any ought model $\mathcal{M}$, if $\mathcal{M}, m/h \models [\alpha$ dstit: $A]$, then $\mathcal{M}, m/h' \models [\alpha$ dstit: $A]$ for each $h' \in \mathit{Choice}^m_\alpha(h)$; and likewise for an indexed ought model $\mathcal{M}$, if $\mathcal{M}, m/h \models_i [\alpha$ dstit: $A]$, then $\mathcal{M}, m/h' \models_i [\alpha$ dstit: $A]$ for each $h' \in \mathit{Choice}^m_\alpha(h)$.*

Lemma 3 *Let $\mathcal{M}$ be an indexed ought model in which the set $\mathit{Ought}_\alpha(m)$ is defined for each agent α and moment m, and let the set $\mathit{Ought}^*_\alpha(m)$ be defined as in steps (1) through (4) above. Then for each choice cell K belonging to the partition Choice^m_α, the set $K \cap \mathit{Ought}_\alpha(m) \neq \emptyset$ iff the set $K \cap \mathit{Ought}^*_\alpha(m) \neq \emptyset$.*

Proof (sketch). This proof relies on the assumption—known as the assumption of *independence of agents*, and not discussed in the present paper—that simultaneous actions by distinct agents are independent: at each moment, for any way of selecting one choice cell from each agent's choice partition, the intersection of all the selected choice cells must be nonempty. More formally, this independence assumption tells us that, whenever s is a function from $Agent$ into subsets of $H_{(m)}$ satisfying the condition that $s(\alpha) \in Choice_\alpha^m$, we have

$$\bigcap_{\alpha \in Agent} s(\alpha) \neq \emptyset.$$

For each moment m, let us define Σ_m as the set of functions f from agents into subsets of $H_{(m)}$ subject to the condition that, for each $\alpha \in Agent$,

$$f(\alpha) \in Choice_\alpha^m \text{ and } f(\alpha) \cap Ought_\alpha(m) \neq \emptyset.$$

We point out two features of Σ_m. First, each function contained in this set satisfies the antecedent conditions in the assumption of independence of agents; and so this assumption allows us to conclude, for each $f \in \Sigma_m$, that

$$\bigcap_{\alpha \in Agent} f(\alpha) \neq \emptyset.$$

Second, Σ_m can be used to characterize the set $Ought_\alpha^*(m)$. For as we recall from step (4) above,

$$Ought_\alpha^*(m) = \bigcap_{\alpha \in Agent} \bigcup \{K : K \in Choice_\alpha^m \text{ and } K \cap Ought_\alpha \neq \emptyset\};$$

and so set theoretic reasoning (involving the axiom of choice for the infinite case) allows us to conclude that

$$Ought_\alpha^*(m) = \bigcup_{f \in \Sigma_m} \bigcap_{\alpha \in Agent} f(\alpha).$$

Using this new characterization of $Ought_\alpha^*(m)$, we now verify the lemma from left to right; the other direction is similar.

Suppose that $K \cap Ought_\alpha(m) \neq \emptyset$ for some $K \in Choice_\alpha^m$. Evidently, the set $K \cap Ought_\alpha^*(m)$ can now be characterized as

$$K \cap \left[\bigcup_{f \in \Sigma_m} \bigcap_{\alpha \in Agent} f(\alpha) \right],$$

which is identical to

$$\bigcup_{f \in \Sigma_m} \left[K \cap \bigcap_{\alpha \in Agent} f(\alpha) \right].$$

It is clear from the definition of Σ_m and the conditions on K, that $f(\alpha) = K$ for some function $f \in \Sigma_m$; and of course, for this particular function f, we have

$$\left[K \cap \bigcap_{\alpha \in Agent} f(\alpha) \right] = \bigcap_{\alpha \in Agent} f(\alpha).$$

Since, as we recall, $\bigcap_{\alpha \in Agent} f(\alpha) \neq \emptyset$, and since

$$\left[K \cap \bigcap_{\alpha \in Agent} f(\alpha) \right] \subseteq K \cap Ought^*_\alpha(m),$$

we can now conclude that that $K \cap Ought^*_\alpha(m) \neq \emptyset$. ∎

It should be noted that the problem this proof overcomes is the fact that, although each of $K \cap Ought_\alpha(m)$ and $K \cap Ought^*_\alpha(m)$ must be nonempty if the other is, these two sets will not in general contain the same histories, and indeed, might not even intersect.

At this point, we are able to complete the proof of Theorem 1 by establishing:

Lemma 4 *Let $\mathcal{M}$ be an indexed ought model, and let $[\mathcal{M}^*]_i$ be defined from $\mathcal{M}$ as in steps (1) through (4) above. Then for each m/h pair and for each sentence A, we have $\mathcal{M}, m/h \models_i A$ iff $[\mathcal{M}^*]_i, m/h \models_i A$.*

Proof. Again by induction on the complexity of the formula A; and again, the only interesting case is that in which A has the form $\bigcirc[\alpha\ dstit{:}\ B]$. We verify the lemma from left to right for this case; the other direction is similar.

Suppose, then, that $\mathcal{M}, m/h \models_i \bigcirc[\alpha\ dstit{:}\ B]$—that is,

(†) for each $h' \in Ought_\alpha(m)$ we have $\mathcal{M}, m/h' \models_i [\alpha\ dstit{:}\ B]$.

We wish to show that $[\mathcal{M}^*]_i, m/h \models_i \bigcirc[\alpha\ dstit{:}\ B]$, or that

(‡) for each $h' \in Ought^*_\alpha(m)$ we have $[\mathcal{M}^*]_i, m/h' \models_i [\alpha\ dstit{:}\ B]$.

So pick some history $h' \in Ought^*_\alpha(m)$. Of course, h' belongs to the particular choice cell $Choice^m_\alpha(h')$; and so $Choice^m_\alpha(h') \cap Ought^*_\alpha(m) \neq \emptyset$. Lemma 3 then tells us that

$Choice^m_\alpha(h') \cap Ought_\alpha(m) \neq \emptyset$ as well—that is, there must be some history h'' belonging to the set $Choice^m_\alpha(h') \cap Ought_\alpha(m)$. By (†) we thus know that $\mathcal{M}, m/h'' \models_i [\alpha\ dstit{:}\ B]$; and then, since of course $h' \in Choice^m_\alpha(h'')$, Lemma 2 tells us that $\mathcal{M}, m/h' \models_i [\alpha\ dstit{:}\ B]$. The formula $[\alpha\ dstit{:}\ B]$ is simpler than $\bigcirc[\alpha\ dstit{:}\ B]$, and so the hypothesis of induction now allows us to conclude that $[\mathcal{M}^*]_i, m/h' \models_i [\alpha\ dstit{:}\ B]$. Therefore, since h' was chosen arbitrarily, this suffices to establish (‡). ∎

Acknowledgments

We are both grateful for the helpful comments of Annette Baier, Paul Bartha, Michael Slote, and an anonymous referee provided by this journal. In preparing the paper, Horty was aided by research support from the National Endowment for Humanities through a Fellowship for University Teachers.

References

[1] Lennart Åqvist. Deontic logic. In Dov Gabbay and Franz Guethner, editors, *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, pages 605–714. D. Reidel Publishing Company, 1984.

[2] Paul Bartha. Conditional obligation, deontic paradoxes, and the logic of agency. *Annals of Mathematics and Artificial Intelligence*, forthcoming.

[3] Nuel Belnap. Backwards and forwards in the modal logic of agency. *Philosophy and Phenomenological Research*, 51:8–37, 1991.

[4] Nuel Belnap. Before refraining: concepts for agency. *Erkenntnis*, 34:137–169, 1991.

[5] Nuel Belnap and Paul Bartha. Marcus and the problem of nested deontic modalities. Manuscript, Philosophy Department, University of Pittsburgh, 1993.

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

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

[8] Nuel Belnap and Michael Perloff. The way of the agent. *Studia Logica*, 51:463–484, 1992.

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

[10] Mark Brown. Action and ability. *Journal of Philosophical Logic*, 19:95–114, 1990.

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

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

[13] Brian Chellas. Time and modality in the logic of agency. *Studia Logica*, 51:485–517, 1992.

[14] Roderick Chisholm. Supererogation and offence: a conceptual scheme for ethics. *Ratio*, 5:1–14, 1963.

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

[16] Roderick Chisholm. He could have done otherwise. *Journal of Philosophy*, 64:409–417, 1967.

[17] Michael Dummett and E. J. Lemmon. Modal logics between S4 and S5. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 5:250–264, 1959.

[18] Fred Feldman. *Doing the Best We Can: An Essay in Informal Deontic Logic*. D. Reidel Publishing Company, 1986.

[19] Dagfinn Føllesdal and Risto Hilpinen. Deontic logic: an introduction. In *Deontic Logic: Introductory and Systematic Readings*, pages 1–35. D. Reidel Publishing Company, 1971.

[20] Harry Frankfurt. Alternate possibilities and moral responsibility. *Journal of Philosophy*, 66:828–839, 1969.

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

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

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

[24] John Horty. An alternative stit operator. Manuscript, Philosophy Department, University of Maryland, 1989.

[25] John Horty. Agency, obligation, act utilitariansim, and coordination. Manuscript, Philosophy Department and Institute for Advanced Computer Studies, University of Maryland, 1993.

[26] John Horty. Deontic logic as founded on nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence*, forthcoming.

[27] I. L. Humberstone. Two sorts of 'ought's. *Analysis*, 32:8–11, 1971.

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

[29] 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.

[30] Anthony Kenny. *Aristotle's Theory of the Will*. Yale University Press, 1979.

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

[32] Michael Perloff. Stit and the language of agency. *Synthese*, 86:379–408, 1991.

[33] Michael Perloff. On the logical grammar of imperatives. Manuscript, Philosophy Department, University of Pittsburgh, 1993.

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

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

[36] Krister Segerberg. Getting started: beginnings in the logic of action. Manuscript, Philosophy Department, University of Auckland, 1989.

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

[38] Richmond Thomason. Deontic logic and the role of freedom in moral deliberation. In Risto Hilpinen, editor, *New Studies in Deontic Logic*, pages 177–186. D. Reidel Publishing Company, 1981.

[39] 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.

[40] 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.

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

[42] P. van Inwagen. Ability and responsibility. *Philosophical Review*, 87:201–224, 1978.

[43] Franz von Kutschera. Bewirken. *Erkenntnis*, 24:253–281, 1986.

[44] Georg Henrik von Wright. *Norm and Action: A Logical Enquiry.* Routledge and Kegan Paul, 1963.

[45] Georg Henrik von Wright. Replies. In Juha Manninen and Raimo Tuomela, editors, *Essays on Explanation and Understanding: Studies in the Foundations of Humanities and Social Sciences*, pages 371–413. D. Reidel Publishing Company, 1976.

[46] Georg Henrik von Wright. On the logic of norms and actions. In Risto Hilpinen, editor, *New Studies in Deontic Logic*, pages 3–35. D. Reidel Publishing Company, 1981.

[47] Ming Xu. Logics of deliberative stit. Manuscript, Philosophy Department, University of Pittsburgh, 1992.