

## International Phenomenological Society

---

Backwards and Forwards in the Modal Logic of Agency

Author(s): Nuel Belnap

Source: *Philosophy and Phenomenological Research*, Vol. 51, No. 4 (Dec., 1991), pp. 777-807

Published by: International Phenomenological Society

Stable URL: <http://www.jstor.org/stable/2108182>

Accessed: 26/04/2010 15:22

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=ips>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



*International Phenomenological Society* is collaborating with JSTOR to digitize, preserve and extend access to *Philosophy and Phenomenological Research*.

# Backwards and Forwards in the Modal Logic of Agency

NUEL BELNAP

*University of Pittsburgh*

By a “modal” logic I mean a logic whose grammar includes an intensional construct having sentences as both inputs and outputs. A “modal logic of agency” intends that some such construct express agency (or action), as, for example, the English construct “ $\alpha$  sees to it that  $Q$ .” In the following four sections I (1) give a brief retrospective, (2) restate some of the claims of Belnap and Perloff 1988, (3) motivate the desirability of a modal logic of agency, and (4) draw ten pictures illustrating some of the cases in which such a logic can provide clarification.<sup>1</sup>

## 1 Mini-history

In this section I give an edge-of-the-thumbail sketch of what I know of the history of the modal logic of agency up to 1989.<sup>2</sup>

The modal logic of agency is not popular. Perhaps largely due to the influence of Davidson (see the essays in Davidson 1980), but based also on the very different work of such as Goldman 1970 and Thomson 1977, the dominant logical template takes an agent as a wart on the skin of an action, and takes an action as a kind of event. This “actions as events” picture is all ontology, not modality, and indeed, in the case of Davidson, is driven by the sort of commitment to first-order logic that counts modalities as Bad. The

---

<sup>1</sup> An early version of this paper was read at the 1989 meeting of the Pacific Division of the American Philosophical Association as part of a symposium entitled “A retrospective of modal logic,” and a later version that was circulated in April, 1989, drew helpful comments from P. Bartha, D. Davidson, D. Elgesem, W. G. Lycan, and D. Makinson, and fresh information from Elgesem (unpublished work and references to Åqvist, Segerberg, and von Kutschera), I. Humberstone (unpublished work including reference to Anselm), D. P. Henry, and D. Walton. Thanks are due M. Perloff for countless suggestions at every stage.

<sup>2</sup> See Segerberg 1989 for an historical account that is much more informative on the topics it treats than is this one. Segerberg’s authoritative piece describes with a deft combination of perspective and detail the contributions of Anselm, von Wright, Fitch, Kanger, Chellas, and Pratt.

project has had some successes, all of which I shall ignore, and some failures, most of which I shall ignore.<sup>3</sup> Certain of its failures, however, are to be attributed to the neglect of the modal features of agency.

I have in mind the tendency of the program to play down the question of how agentive constructions embed in larger constructions. This tendency is explicit in the initial description of Davidson 1967 of its own task: "I would like to give an account of the logical or grammatical role of the parts or words of [simple sentences about actions] that is consistent with the entailment relations between such sentences... ."

What is from the modal point of view striking is that the aim set out in this passage includes only *half* of what is needed for a compositional account of meaning. Davidson 1967 sets out to show how "the meanings of action sentences depend on their structure," but does *not* begin with the aim of showing how the meanings of sentences that *contain* action sentences depend on *their* structure. The stated aim does not include, for instance, telling how the meaning of "Jones refrained from buttering the toast" or "Mary demanded that Jones butter the toast" or "Jones, butter the toast!" or "How speedily did Jones butter the toast?" or "Jones brought it about (or saw to it) that Jones buttered the toast" depend on the meaning of "Jones buttered the toast," or perhaps telling how they don't if they don't. Half the compositional problem has been left out of the initial statement of purpose.

The modal logic of agency should strike the other way. The modal logic of agency should be interested in larger contexts *containing* agentive sentences.

The fact is that with regard to embedding agentive sentences in larger contexts, it makes a difference that they are agentive, and it makes a difference who the agent is. Embedding contexts care about these things. The reason that the action-as-event paradigm has not contributed to our understanding of the embedding of agentives is, perhaps, that its resources do not permit it to do so. Here is a slogan that gives a smallish part of the explanation why: propositions and sentences have negations, but actions and events do not.

The following observation may confirm this conjecture. At the end of Davidson 1967 the question is raised of the intentionality of action. With clear recognition that what is being addressed is a part of the other half of the compositional problem, it is there proposed that we introduce intention by means of an explicitly embedding expression such as

it was intentional of  $x$  that  $p$ , where ' $x$ ' names the agent and ' $p$ ' is a sentence that says the agent did something.

---

<sup>3</sup> See Bennett 1988 for an indispensable perspective.

Although the proposal presumably requires for its coherence that  $p$  display a term for “the agent” in some recoverable fashion, it seems to borrow no other feature from the earlier thesis that the logical form of the embedded sentence involves an ontology of events or actions. The indicated lack of influence of the earlier thesis on the later proposal is contrary to expectations, since generally in compositional semantics a view about the logical form of a “part” of a certain kind both constrains and is constrained by a view about the logical form of an expression that embeds just that kind of part. I think that this is another indication of the difficulty of using the picture of actions-as-events as a guide in understanding the role of agentive sentences when they are embedded in larger contexts.

The earliest treatment of the modal logic of agency of which I have learned is that of Anselm of Canterbury about 1100. In the document that Henry 1967 calls  $N$ ,<sup>4</sup> Anselm writes

Quidquid autem ‘facere’ dicitur, aut facit ut sit aliquid, aut facit ut non sit aliquid. Omne igitur ‘facere’ dici potest aut ‘facere esse’ aut ‘facere non esse’ (p. 124; from  $N$  29.8.10).

Paraphrase by Henry: For all  $x$ , if ‘ $x$  does’ is true, then  $x$  does so that something either is so or is not so. Hence the analysis of ‘doing’ will in fact be an analysis of  $x$ ’s doing so that  $p$ , and of  $x$ ’s doing so that not- $p$  [where ‘ $p$ ’ is a clause describing a state of affairs, and ‘not- $p$ ’ is short for ‘it is not the case that  $p$ ’] (p. 124).

Anselm goes on to describe a kind of square of opposition that clearly indicates he had in mind a modal logic of agency (to the extent to which that can be said without anachronism), but his work seems to have remained unnoticed until after the stirring of modal logic in this century.<sup>5</sup>

The first modern desire for a modal agentive construction seems to have been felt by philosophers working their various ways through the embedding requirements of legal and deontic concepts. One can certainly see the need expressed in the pioneering work of Hohfeld 1919, though there the agency construction is always embedded in additional constructions imputing legal rights, duties, powers, etc., as in locutions such as the following on p. 38:

$X$  has a right against  $Y$  that he shall stay off the former’s land.

---

<sup>4</sup> The date of *circa* 1100 for  $N$  is implied by the sources available to me around the house, i.e., the discussion on p. 120 of Henry 1967 together with the “Anselm” entry in the *Encyclopedia Britannica* of 1968.

<sup>5</sup> If you promise to accept my remark as merely helpful rather than authoritative, I will hazard that Henry 1953 is the first reference to Anselm that appreciates his work as modal, and that Chisholm 1964a (who cites Henry 1960) is the earliest reference to Anselm by an active researcher in this field. Other references to Anselm on this topic: Danto 1973, Humberstone 1976 (the reference does not appear in the published abstract Humberstone 1977), the perceptive Walton 1976, 1976a, and 1980 (which cites Dazeley and Gombocz 1979), and a sterling account in Segerberg 1989.

The next place I know it to crop up, much more explicitly but still embedded in the context of a normative expression, is in Kanger 1957:

Ought(*Y* sees to it that  $F(X, Y)$ ) (p. 42).

Although the locution “sees to it that” is displayed only in a normative context and wholly without comment, it is clear from the general tenor of Kanger’s methodology that he intended to be isolating a norm-free concept of agency.

The explicit grammatical breakthrough for the logic of agentive modality comes in Anderson 1962, who, reflecting on Hohfeld,<sup>6</sup> for the first time introduces a stand-alone form of expression intended to disengage the concept of agency from normative considerations. When on p. 40 Anderson takes

$M(x, p, y)$

to represent the case “when *x* executes what is regarded as an ‘action’ ... and *y* is the recipient or patient of the action executed by *x*,” he all of a sudden gives us a clean target for some analytic questions that otherwise come out all confused.

Anderson sometimes reads  $M(x, p, y)$ , with perhaps too little attention to the connections between formal and English grammar, as “*x* does *p* to (for) *y*.” Evidently here agency is, for better or worse, not separated from patiency. And certainly there is in Anderson no semantic theory of agency or patiency, and only a trace of a deductive calculus (e.g. on pp. 45–46 Anderson points out that the implication between  $\sim M(x, p, y)$  and  $M(x, \sim p, y)$  goes only from right to left). That is, Anderson pioneers in isolating agency and patiency, but he does so only immediately to recombine them with deontic concepts.

In 1963 each of two logicians, Fitch and von Wright, advanced modal theories of agency, each of them stressing syntactic developments. Fitch 1963 defines “does *A*” in terms of two other modalities, “striving for” and “causes,” and offers a deductive calculus. The work has not been taken up by later logicians and is seldom cited in the published literature.<sup>7</sup>

Von Wright, beginning with von Wright 1963, and continuing at least through von Wright 1981, was I think the other logician to be a first to treat

---

<sup>6</sup> Very likely after correspondence or conversations with his friend, Kanger. Somewhat later Anderson visited Manchester, where Henry was. Henry remarked in personal correspondence that during this year of 1965 there was a colloquium involving a number of persons interested in agency, including e.g. Hare and Kenny.

<sup>7</sup> Indeed, although I was Fitch’s admiring and fond student and colleague, I regret to say that I had to be reminded of this paper by Segerberg 1989, which contains a maximally useful account. It is a pleasure tinged with sadness to add that it was certainly Fitch and Anderson who imprinted on me the possibilities in and importance of a modal logic of agency.

agency (or action) as a specific modal or quasi-modal topic, always with that specially honest von Wrightian insistence of the lack of finality of the formulation in question, including attending to non-modal formulations in which complements are taken as terms signifying specific or generic actions, rather than sentences. As in other cases, the work keeps a close eye on deontic logic, to which he contributed so much. Von Wright did not I think succeed in disentangling agency from change, and did not evince interest in the general problem of embedding of agentives. For instance (but only ‘for instance’), in his earliest paper von Wright took as a primitive

$d(p/p)$ ,

to be read as expressing some such idea as “the agent preserves the state described by  $p$ ” (pp. 43, 57).<sup>8</sup> In contrast with Anderson, however, agency here has been separated out from patiency.

Kanger and Kanger 1966 introduce as a separate locution

$X$  causes  $F$ ,

where  $F$  is supposed to be a sentence, but in a fashion like Anderson’s, they logicize about it only by setting down that  $F$  may be replaced by its logical equivalents, and that the proposition that  $X$  causes  $F$  implies that  $F$ .

Three influential lines of research began about the same time, each of which highlighted the separate existence of agentive modalities; namely, those initiated by Castañeda, by Kenny, and by Chisholm.<sup>9</sup>

Castañeda, whose views concerning deontic logic have informed both philosophers and logicians for many years (since at least Castañeda 1954), has much to say that is relevant to agency as a modality. Though his philosophical concerns have led him to pursue goals other than the formulation of a modal logic of agency, he has repeatedly urged the fundamental importance of the grammatical and logical distinction between “propositions” and “practitions” (a distinction put as clearly as anywhere in Castañeda 1981); but because there is no possibility of constructing a Castañeda “practition” from an arbitrary sentence, in the way for instance that Anderson’s  $M(x, p, y)$  or von Wright’s  $d(p/p)$  each permit an arbitrary sentence in place of  $p$ , Castañeda practitions cannot themselves serve as the foundation for such a modal logic of agency.

---

<sup>8</sup> Von Wright tends to leave to the reader the task of putting bits of logical grammar together with bits of English grammar.

<sup>9</sup> Of course other work on the theory of action has also influenced the modal logic of agency, but that literature is unsurveyably vast. I note as a passing example that there is hardly a one of my past or present departmental colleagues who has not contributed.

Kenny 1963, in the course of initiating a rich literature on the verbal structure of our causal and agentive discourse, says that any “performance” in his technical sense is describable in the form

bringing it about that  $p$ .

And Chisholm 1964 takes the following as a basic locution on which to found an extensive series of definitions and explanations in the vicinity of agency:

There is a state of affairs  $A$  and a state of affairs  $B$ , such that he makes  $B$  happen with an end to making  $A$  happen,

where the letters stand in for “propositional clauses,” and where the subject of “makes happen” can be either a person or a state of affairs. The discussions of Kenny and Chisholm, though relevant to logical questions, are themselves not directed toward the formulation of either proof-theoretical or semantic principles governing their respective basic locutions.

The above is as accurate a record as I can manage of the early history of the modal logic of agency. If this story is right, then the following is its gist.

### History of the modal logic of agency prior to 1969

- *Anselm circa 1100*                      *facere esse* ( $x$  does so that  $p$ )
- *Hohfeld 1919*                               $X$  has a right against  $Y$  that he shall stay off the former’s land
- *Kanger 1957*                                *Ought*( $Y$  sees to it that  $F(X, Y)$ )
- *Anderson 1962*                           $M(x, p, y)$  ( $x$  does  $p$  to [or for]  $y$ )
- *Fitch 1963*                                 Does  $A$
- *Von Wright 1963*                         $d(p/p)$  (the agent preserves the state described by  $p$ )
- *Kanger & Kanger 1966*                 $X$  causes  $F$
- *Castañeda 1954ff, Kenny 1963, Chisholm 1964ff*:    relevant discussions

The first modal logic of agency with an explicit semantics is I think that of Chellas 1969. The primitive locution is

$\Delta\tau\phi$ ,

to be read as “ $\tau$  sees to it that  $\phi$ ,” where  $\tau$  is an agent and  $\phi$  takes the place of a sentence. Chellas only deploys this locution in one context, namely, as the argument of an imperative. (But Chellas does not *restrict* the complement of

an imperative operator to sentences having the form  $\Delta\tau\phi$  as is required by the Restricted Complement Thesis stated below.)

As for semantics, Chellas takes as a paradigm the technique made famous by Kripke not long before Chellas was writing; I mean deployment of a binary relation between “worlds” in order to clarify modal concepts. Chellas in particular gives a semantic clause for  $\Delta\tau\phi$ :

$\Delta\tau\phi$  is true at the present world just in case  $\phi$  is true at all those worlds under the control of—or responsive to the action of—the individual which is the value of  $\tau$  at the present world.

The language that Chellas uses in this pioneering explanation, like the “relative possibility” language of Kripke a few years earlier, is neither familiar in itself nor further clarified by Chellas. Perhaps this is the reason that, like his predecessors, Chellas in practice confines his agentive locution to the imperative context from which his need for it sprung, and does not pause to investigate its separate properties.

After Chellas there is a substantial group of logicians all of whom have deployed a binary relation or a pair of binary relations in an effort to generate a semantic understanding of an agentive modality that might be used as the complement of an imperative or of a deontic operator; I know of Pörn 1970, 1971, 1974, 1977; Needham 1971; Åqvist 1972; Kanger 1972; Hilpinen 1973; Humberstone 1977; Lindahl 1977; and Talja 1980. For a critic on the line of research being described, with special reference to Pörn 1970, see Walton 1975; also of note are Walton 1976, 1976a, 1980, which develop some insights in an independent and more non-semantic fashion.

The earlier Pörn papers and that of Åqvist use only a single binary relation; the idea of using two binary relations seems to be independently due to Needham 1971, Kanger 1972 and Hilpinen 1973.<sup>10</sup> The reason for the second binary relation is given as this: agency has not only a *sufficient* condition aspect but a *necessary* condition aspect (Kanger, p. 109; Hilpinen, p. 119), and one needs a separate relation for each. The later workers in this mini-tradition play variations on this theme. In my judgment this line of investigation, although initially promising, and although producing some useful insights, has not been much followed up for the following reason: it has *remained* obscure what one is to make of the binary relations that serve as the founding elements of the entire enterprise. Kanger 1972 says, for example, that one of the relations holds between a person and a couple of worlds or indices when everything the person does in the second world is the case in the first; and the other relation holds when the opposite of everything the person does in the second is the case in the first (p. 109). That is far from clear, and

---

<sup>10</sup> Unless I have overlooked it, there is no cross-mention. I have not seen Needham's M.A. thesis, but make the inference from Pörn 1977.



no one in the tradition is, in my judgment, any clearer than that. For a final example, I describe and quote at length from Pörn 1977, which among those mentioned above is the most developed grammatical and semantic treatment of agentive modality. (All words not inside quotation marks are mine.)

$D_a p$  is read “it is necessary for something which  $a$  does that  $p$ ” (p. 4). It is said that an equivalent concept is found as the definition of “ $a$  sees to it that  $p$ ” in Chellas 1969, Chapter III, Section 4, and in Pörn 1971.

“... consider all those hypothetical situations  $u'$  in which the agent does at least as much as he does in  $u$ . If  $v$  is such a situation, it may be said to be possible relative to what the agent does in  $u$ . ... if  $p$  is necessary for something that  $a$  does in  $u$ , then there cannot be a situation which is possible relative to what  $a$  does in  $u$  and which lacks the state of affairs that  $p$ . ... A natural minimal assumption is that the relation [of relative possibility] is reflexive and transitive ... .”

$D'_a p$  is read “but for  $a$ 's action it would not be the case that  $p$ ” (p. 5), and also “ $p$  is dependent on  $a$ 's action.”

“... for the articulation of the truth of  $D'_a p$  at  $u$  we require all hypothetical situations  $u'$  such that the opposite of everything that  $a$  does in  $u$  is the case in  $u'$  ... [the relation must be] irreflexive and serial ... .”

Further, to connect the two modalities  $D$  and  $D'$ , a condition is imposed that “requires that worlds which are alternatives to a given world under the relation [for  $D_a$ ] be treated as equals in contexts of a counteraction conditionality.”

$C'_a p$  is read “ $p$  is not independent of  $a$ 's action” (p. 7).

$E_a p$  is defined as the conjunction of  $D_a p$  and  $C'_a p$ , and read “ $a$  brings it about (causes it to be the case that, effects that)  $p$ ” (p. 7). It is said that an equivalent concept is found in Needham 1971, p. 154, an essentially equivalent concept in Hilpinen 1973, Section VI, and explicitly in Pörn 1974, p. 96.

An alternate  $E^*_a p$ , defined as the conjunction of  $D_a p$  and  $D'_a p$ , is found unacceptable. It is said that an equivalent definition is that of Kanger 1972, p. 108.

My point is certainly not that Pörn 1977 is less clear than it can be; quite the contrary, it seems to me to offer the best explanations of and the most detailed working out of the modal logic of agency as based on binary rela-

tional semantics. The proper conclusion is rather that one should doubt the likelihood that the semantics themselves can serve in the way that was hoped.

Åqvist 1974, 1978 provide a much more intuitive semantic setting; these papers are the first of which I know that make the fundamental suggestion that agency is illuminated by seeing it in terms of a *tree structure* such as is familiar from the extensive form of a game as described in von Neumann and Morgenstern 1944. Åqvist's account of agency is in some respects akin to that described below, in some respects less flexible, and in some respects richer. His aim is not strictly to provide a modal logic of agency; for example, the primitive of Åqvist 1978 is "DO( $a, Pa$ )," to be read " $a$  does, or acts, in such a way that he  $Ps$ ," and where " $Pa$ " must be an *atomic* sentence (rather than an arbitrary sentence), and like von Wright, Åqvist wraps agency together with change. But his goal is close enough to warrant (1) a comparison (which is not attempted here) and (2) a suggestion that the reader consult these sources. A notable relevant paper is Mullock 1988. Of decisive importance is the uncommonly rich joint work Åqvist and Mullock 1989, which applies insights derived from the tree structure to serious questions in the law. This book, like that of Hamblin, mentioned below, is *required reading*.

There is one later commentator on the tradition just described who is of special excellence and interest: Makinson 1986.

In a series of more than a dozen papers beginning with Segerberg 1980, and including among others Segerberg 1981, 1982, 1984, 1985, 1985a, 1987, 1988, 1988a, and 1989, a distinguished modal logician develops a richly motivated and intuitively based formal approach to action by taking a *routine* as the guiding concept. Segerberg explicitly bases some of the intuitive and formal aspects of his work on studies that in computer science have come to be called "dynamic logic," the influence on Segerberg being primarily through Pratt. Consult Elgesem 1989 for a sympathetic yet critical penetrating account of Segerberg's line of research. The work is not fully in the *modal* logic of agency, since it stresses a grammar of (1) terms (including complex terms) for naming "actions" and (2) predicates for expressing properties of "actions," and thus self-consciously avoids a grammar of nesting connectives. But instead of a complaint this is intended only as a reason for limiting myself to a mere mention of what may indeed turn out to be not only valuable in itself but a useful link between the ontological and modal points of view on agency.

Mention of Pratt calls attention to the existence of a large and interesting formal literature that I fail to cite as part of this mini-history except in so far as it has influenced Segerberg, namely the work on "dynamic logic" and its cousins that has been done by Floyd, Hoare, Pratt, and other computer scientists (see Segerberg 1989 for brief entrée via Pratt that is written espe-

cially with the logic of action in view, and see Pratt 1980 for an excellent fuller account). There are three reasons for excluding this line of investigation from the present survey: (1) I am very far from familiar with the literature, so that making it accessible is best left to someone else. Further, what I know of it (2) stresses the ontological rather than the modal approach, whereas the latter is the topic of this mini-history, and (3) what I know of it is relevant to action only in the wide sense of “action” that encompasses *mechanical* action, i.e., the sense of “action” that encompasses the action of programs and starter motors. In fact the present modal point of view makes it arguable that this literature is no more relevant to *agency* than is the literature of any other discipline that gives us ways to fill out the sentential complement of “sees to it that”: an agent can see to it that the starter engages and passes through various stages, or that a certain recursive program runs, or ... . But it seems best to make explicit my failure to more than barely mention such a large literature just because so many persons think that although it may be *arguable*, it certainly isn’t *plausible* that it has no special relevance to agency.<sup>11</sup>

Penultimately there is von Kutschera 1986, which articulates in one form or another nearly all of the essential underlying ideas concerning agency on which we base the semantics offered below.<sup>12</sup>

Finally there is Hamblin 1987, which in the context of a study of imperatives provides a rich source of formal, informal, and semi-formal ideas on the topic at hand, many of which have influenced the present work; in particular, collegial reflection on Hamblin’s “action-state semantics” was the immediate context of the beginning of the research reported in the rest of this paper. My own recommendation is that no one ought to try to move deeply into any part of the theory of agency without reading this important book.

## 2 Who wants a logic of *stit* (sees to it that)?

All that work on the modal logic of agency may seem too abstract. Put it this way: who cares about *sees to it that* as over against vigorous verbs such as *butters*? The stakes are considerably upped, however, if you agree to the following claims, all of which have been argued by Belnap and Perloff 1988.

---

<sup>11</sup> On the other hand, I once asked a well-known computer scientist/mathematician after a lecture on parallel processing if he had meant his use of “actor” and “agent” to be anything but an idle metaphor; he was aghast that I should need to inquire.

<sup>12</sup> At the very least, von Kutschera 1986 is to be credited with the No Choice between Undivided Histories condition, with generalization beyond the discrete, with generalization to multiple independent agents (including the Something Happens condition), with attention to strategies, and with semantics for the “deliberative *stit*” mentioned below in note 17. It also needs to be remarked that von Kutschera 1986 cites the earlier von Kutschera 1980.

In order to state them conveniently, let us agree to use *stii* as an abbreviation for *sees to it that*.

The first claim is

*The Stit Complement Thesis.* [ $\alpha$  *stii*:  $Q$ ] is grammatical and meaningful (though perhaps silly) for an *arbitrary* sentence  $Q$ .

The tradition of agentive modal logic described above has done very little in the way of exploiting this thesis, which implies that we ought to *look for* ways of finding sentential complements for *stii*. In particular, the complement might itself either be agentive or not. There is “Ophelia saw to it that she got herself to a nunnery” (in which the complement is itself an agentive) and there is “Ophelia saw to it that she had flowers in her hair” (in which it is not).

The second claim needs statement but surely neither argument nor illustration:

*The Agentiveness of Stit Thesis.* [ $\alpha$  *stii*:  $Q$ ] is always agentive in  $\alpha$ , regardless of its complement.

The third claim comes to this: that a sentence can be paraphrased with  $\alpha$  *stii* is an excellent criterion of agency.

*The Stit Paraphrase Thesis.*  $Q$  is agentive in  $\alpha$  just in case  $Q$  is paraphrasable as (or is strongly equivalent to) [ $\alpha$  *stii*:  $Q$ ]:

$$Q \leftrightarrow [\alpha \textit{ stii}: Q].$$

The fourth claim is that a variety of constructions of great interest to philosophers must take agentives as their complements, and that this remains true even when the complements appear not to be agentives.

*The Restricted Complement Thesis.*

- The imperative construction must take an agentive as its complement.
- Deontic constructions such as obligation, prohibition, and permission must take agentives as their complements.<sup>13</sup>

---

<sup>13</sup> Deontic logicians sometimes study an “impersonal ought,” sometimes called “the ought-to-be.” Some people think there is no such thing as the ought-to-be, and they are probably right; in any case, I mean to be speaking of what would be called the “ought-to-do.”

- Constructions involving ability, power, could-have-done, etc. must take agentives as their complements.
- Miscellaneous other constructions such as those involving requests, invitations, demands, plans, intentions, promises, commitments to do, and lots more, must take agentives as their complements.

It is the combination of these claims that exhibits the potential power of the modal logic of agency, or, as we might as well begin to call it, the logic of *stit*, for the following is an easy consequence of the Restricted Complement Thesis together with the *Stit* Paraphrase Thesis.

*The Stit Normal Form Thesis.* In the study of all of those concepts represented by constructions that take agentives as complements, nothing except confusion is lost if the complements are taken to be all and only *stit* sentences.

For example, if your theory of obligation addresses every case of *obligated to see to it that*, then your theory will be complete. And if your theory of the obligation construction ever lets it take as complement a sentence that cannot be paraphrased via *stit*, then your theory will be wrong. (In the sequel I will sometimes say that the constructions requiring agentives *must* take *stit* sentences, meaning thereby only that they must take sentences that are paraphrasable as *stit* sentences.)

Given these theses, the desire for a logic of *stit* now begins to fall into place. Let me illustrate by telegraphing just a few examples.<sup>14</sup>

1. Deontic logicians ask if obligation is closed under conjunction. But that is at least superficially an ill-formed question, since the complement of an obligation must (according to the *Stit* Normal Form Thesis) be a *stit* sentence. In seeking a replacement question, we must look to the logic of *stit* to tell us whether or not a conjunction of *stit* sentences is itself equivalent to a *stit* sentence, and if so, to which one. For example, the *stit* normal form guides us to differentiate the following questions.

Is  $[\alpha \textit{ stit}: P] \& [\alpha \textit{ stit}: Q]$  equivalent to  $[\alpha \textit{ stit}: P \& Q]$ ?

Is  $[\alpha \textit{ stit}: P] \& [\alpha \textit{ stit}: Q]$  equivalent to  $[\alpha \textit{ stit}: [\alpha \textit{ stit}: P] \& [\alpha \textit{ stit}: Q]]$ ?

2. Deontic logicians ask if permission distributes over disjunction. But that is an ill-formed question, since the complement of a permission must be a *stit* sentence. In seeking a replacement question, we must look to the

---

<sup>14</sup> In order to leave the reader with something definite to try, the problems numbered 1 and 2 below are not explicitly solved in what follows.

logic of *stit* to tell us whether or not a disjunction of *stit* sentences is itself equivalent to any *stit* sentence.

3. Deontic logicians ask if prohibition is the same as being obligated not to do. But that is an ill-formed question, since the complement of an obligation must be a *stit* sentence. In seeking a replacement question, we must take into account what surely we had better build in to the logic of *stit* sentences from the beginning: generally the negation of a *stit* sentence is *not* equivalent to a *stit* sentence. That is, one can well have  $\sim[\alpha \textit{stit}: Q]$  without there being any  $P$  such that  $\sim[\alpha \textit{stit}: Q] \leftrightarrow [\alpha \textit{stit}: P]$ ; for example, you did not see to it that the sickle moon hung high last night amid the splendid stars, but there is no way in which you *saw* to your “failure” to see to that theatrical state of affairs.
4. Deontic logicians ask if obligation is closed under logical or analytic consequence. But since the complements of the obligation construction are one and all *stit* sentences, one had better first understand consequence among the *stit* sentences themselves, taking into consideration that *stit* itself is surely not closed under consequence: I can see to it that an injured man is bandaged without seeing to it that there is an injured man.
5. A related but not equivalent question is this: is that to which I see closed under *modus ponens* for material implication? Do  $[\alpha \textit{stit}: P \supset Q]$  and  $[\alpha \textit{stit}: P]$  together imply  $[\alpha \textit{stit}: Q]$ ? There is a curious answer to this seemingly obvious question.
6. Metaphysicians worry about “could have done otherwise”, but without the help of *stit*, it is at best confusing to say what “otherwise” means; one runs out of places to put a negation. With *stit*, and keeping in mind that “could have” must take a *stit* complement, an approximation to the topic is sensibly taken to be the following: given  $[\alpha \textit{stit}: Q]$ , does it follow that  $\alpha$  could have seen to it that  $\sim[\alpha \textit{stit}: Q]$ ? For example, given that  $\alpha$  saw to it that his children were educated, does it follow that  $\alpha$  could have seen to it that he did not see to it that his children were educated?

Conclusion: if a well-based logic of *stit* can answer these questions in a satisfying way, then it has a good deal to offer.

### 3 Semantics for *stit*

Let us approach the logic of *stit* semantically, and somewhat indirectly, by going after the idea that could be expressed by saying

the present fact that  $Q$  is guaranteed by a prior choice of  $\alpha$ .

The displayed expression itself is not suggested as an adequate analysis of *stit*, which it obviously is not, but instead as a useful *approximation*. The strategy is to quantify out the prior choice, thereby leaving out any trace of

an intentional relation between  $\alpha$  and the fact that  $Q$  itself, and even leaving open the existence or nature of an intentional element in the prior choice. The strategy of approximation is successful to the extent that the resulting structure attributed to *stit* helps us when we are confused; as Braithwaite says, “no calculus without calculation.”

We construct a semantic picture of *stit* against the backdrop of a picture of *branching time* in the sense of e.g. Thomason 1970 or McCall 1976.<sup>15</sup> The metaphysical backdrop, then, has *moments* ordered by earlier/later into a tree-like structure, with upward branching representing an openness as to the future, and the absence of backward branching representing the settledness of the past, as suggested by Figure 1 just below.

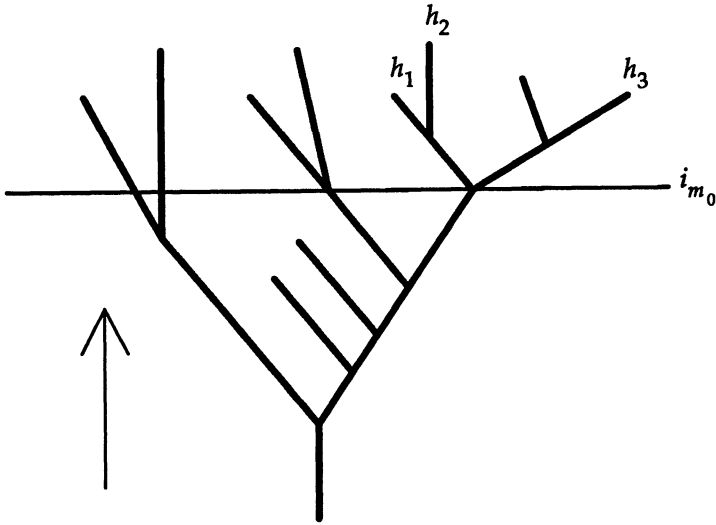


Figure 1: Branching time: moments, histories, and instants

A *history* is a set of moments constituting a single complete branch of the tree. I follow Thomason in holding that truth is fundamentally relative to moment-history pairs; but because of the special nature of *stit*, I can speak in this introductory discussion as if truth were relative only to moments. Let us also suppose that moments can be partitioned “horizontally” into *instants* in such a way that their order is exactly replicated. Suppose also at least for the present discussion that each instant intersects each history at a unique

<sup>15</sup> I am of course not suggesting that the idea of branching time began with these papers; for one thing, it is well known that the contemporary flowering of tense logic in all its ramifications is squarely due to Prior, with branching time to be found in Prior 1967. The reason I single out Thomason 1970 is that I adopt his semantic point of view, and the reason I single out McCall 1976 is that I adopt his stance that what is being presented is objective metaphysics.

moment, and that each two moments are “historically connected” via a common earlier moment (but do not suppose that there is a latest such).<sup>16</sup> Letting  $i_{m_0}$  be the instant determined by  $m_0$ , we may think of its members as alternate ways of “filling” the same instant of time as is “filled” by  $m_0$ . We say that two histories are *undivided* at a moment,  $m_0$ , if at  $m_0$  they appear as if they were a single line. That is, not only must they pass through  $m_0$ , but they must also share some later moment. Example: in Figure 1, histories  $h_1$  and  $h_2$  are undivided at moment  $m_0$ , but histories  $h_1$  and  $h_3$  are not undivided at moment  $m_0$ .

I have announced that the current approximation to *stit* is to be based on the choices of agents; and since one cannot get something for nothing, the branching-time backdrop must be supplemented with agents and their choices, to be placed centerstage.

We postulate that the concept of *agent* is absolute in the sense of Bressan 1972 (or better, a substance sort in the sense of Gupta 1980), which in particular means that there is no fission and no fusion of agents.

As for choice, we idealize by postulating that at each moment,  $w_0$ , there is defined for each agent  $\alpha$  (possibly one-member) *choice set*, that is, a partition of all of the histories passing through  $w_0$ . A member of a choice set is called a *possible choice*, so that a possible choice is a set of histories. If there is only one possible choice for  $\alpha$  at  $w_0$ , it will be the set of *all* histories passing through  $w_0$ .

This is a metaphysical postulate, but it is not wild-eyed, and that for two reasons. The first reason is that it is subject to two sane conditions. The first condition, due to P. Kremer, is the *No Choice between Undivided Histories condition*: a choice set for an agent at  $w_0$  must keep together histories that are undivided at  $w_0$ ; i.e., no agent can make a choice that includes one of two undivided histories but excludes the other. The second condition, the *Something Happens condition*, only comes into play when considering multiple agents: for each way of selecting one possible choice for each agent from among his or her choice set, the intersection of all the possible choices selected must contain at least one history; i.e., something happens. The second reason that our metaphysics of choices is not crazy is that it explicitly allows that the choice set for an agent at a moment might be vacuous, containing but a single option, namely, the set of all histories passing through that moment. Perhaps this is the situation that most of us are in most of the time, and it certainly describes our state when asleep.

---

<sup>16</sup> The idea is that each history is isomorphic to each other under the one-one correspondence provided by the partition into instants. Two further remarks: although we do postulate “historical connection,” we do not here rely on this postulation; and although we fail to postulate that the tree of moments is a lower semi-lattice, we do not here *rely* on this failure.



That's all the metaphysics. Summary: to branching time we add a concept of agent, and a (possibly vacuous) choice set for each agent at each moment. Observe that the settledness of the past implies that if moment  $m_0$  is later than moment  $w_0$ , then there is exactly one possible choice for  $\alpha$  at  $w_0$  containing a history leading to  $m_0$ .

Definition: if histories from later moments  $m_0$  and  $m_1$  pass through the same possible choice for  $\alpha$  at an earlier moment  $w_0$ , and if  $m_0$  and  $m_1$  inhabit the same instant (i.e., if  $i_{m_0}=i_{m_1}$ ), then we say that  $m_0$  and  $m_1$  are *choice-equivalent* for  $\alpha$  at  $w_0$ : no choice that  $\alpha$  can make at  $w_0$  can tell them apart.

We can now give a semantic account of [ $\alpha$  stit:  $Q$ ], where the intension of  $\alpha$  falls under the concept of agent, as follows.<sup>17</sup>

[ $\alpha$  stit:  $Q$ ] is true at  $m_0$  just in case there is a prior "choice point"  $w_0$  satisfying two conditions.

1. *Positive condition.*  $Q$  must be true at all moments that are choice-equivalent to  $m_0$  for  $\alpha$  at  $w_0$ . (Thus, the prior choice of  $\alpha$  at  $w_0$  "guarantees" that  $Q$  holds at the instant  $i_{m_0}$  determined by  $m_0$ .)<sup>18</sup>
2. *Negative condition.* There must be some moment  $m_0'$  that (i) is in the instant  $i_{m_0}$  determined by  $m_0$ , (ii) lies on a history through  $w_0$ , and (iii) is such that  $Q$  is *not* settled true there. (Thus, the choice set for  $\alpha$  at  $w_0$  is *really* a choice with respect to  $Q$ ; it was not the case that  $Q$  would have been guaranteed at the instant  $i_{m_0}$  determined by  $m_0$  independently of the choice of  $\alpha$  at  $w_0$ .)

#### 4 Stit pictures

All this is more easily communicated visually than verbally. The rest of this paper is accordingly devoted to showing some pictures, each one designed to

<sup>17</sup> The metaphysical stage-setting described above permits the introduction of more than a single locution that might be read in English as *stit*. Of fundamental interest is one introduced in discussion by J. Horty (and much earlier in von Kutschera 1986, as observed in note 12): [ $\alpha$  stit:  $Q$ ] is true at a moment-history pair  $(m_0, h_0)$  just in case (i)  $Q$  is true at every moment-history pair  $(m_0, h_1)$ , where  $h$  is drawn from the same possible choice for  $\alpha$  at  $m_0$  as is  $h_0$ , and (ii)  $Q$  is not settled true at  $m_0$ . In this context  $Q$  would typically be future tensed. Sometimes, when both *stits* are under discussion, I call the von Kutschera/Horty one the "deliberative" and the Belnap/Perloff one described in this paper the "achievement" *stit*, and use the notations "*dstit*" and "*astit*" to avoid confusion.

<sup>18</sup> We are grateful to W. Edelberg for pointing out to us that the statement of the Positive condition in Belnap and Perloff 1988 became garbled in our effort to "informalize" (as S. K. Thomason says); please ignore it, relying instead on the statement of this paper.

answer some hovering question. It is in fact the possibility of meaningful pictures, I think, that constitutes the chief advantage of this semantics over those previously described. The pictures are a way to give force to the Braithwaite slogan, “no calculus without calculation.” (So that the pictures can be efficient conceptual supports of calculation, and at the same time convey intuitive significance, please give them the benefit of the doubt by treating as not relevant what is not explicitly drawn.)

### Simple *stit*

Question: what do the positive and negative conditions for *stit* look like in a simple case? Here is the first picture, without much explanation, just to help make vivid the Positive and Negative conditions.

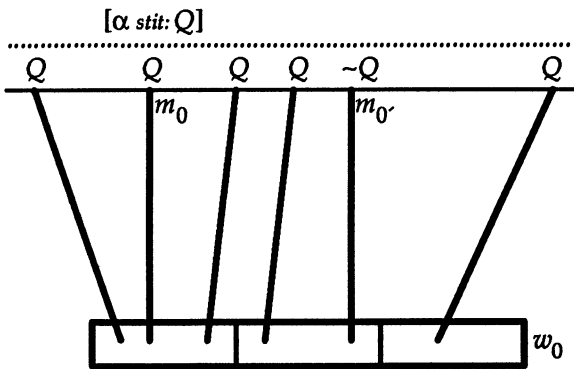


Figure 2: Simple *stit*

The long rectangle is an enlarged picture of a single moment,  $w_0$ . The rectangle is divided into separate boxes, each of which represents a possible choice for  $\alpha$ , so that at  $w_0$ ,  $\alpha$  has three possible choices. But  $\alpha$  does not have control of everything at  $w_0$ ; if  $\alpha$  chooses left, each of three things can happen, and if  $\alpha$  chooses middle, each of two things can happen. So the left choice by  $\alpha$  leaves it undetermined exactly what is to happen; you can see that  $Q$  is nevertheless true at each of the three moments that are choice-equivalent to  $m_0$  (Positive condition), and that  $Q$  is false at some appropriate moment,  $m_0'$ , as well (Negative condition). Therefore,  $w_0$  bears witness to the truth of  $[\alpha \text{ stit}: Q]$  at  $m_0$ . This is explained a little bit more in relation to the next picture.

### Paradigm *stit*

Question: what is a paradigm case of *stit*, in the guise of the approximation “the present fact that  $Q$  is guaranteed by a prior choice of  $\alpha$ ,” and with the semantics suggested for this construction?

The box in Figure 3 (which looks like and is supposed to look like a von Neumann normal form of a game) is a blowup of a single moment,  $w_0$ . Assume that  $\alpha$  and  $\beta$  are the only agents, and further assume that there are no nonvacuous choices for either  $\alpha$  or  $\beta$  at any moment except  $w_0$  (in particular, the forkings in Figure 3 that are pictured *above*  $w_0$  represent possibilities for Nature, not choices for  $\alpha$  or  $\beta$ ).

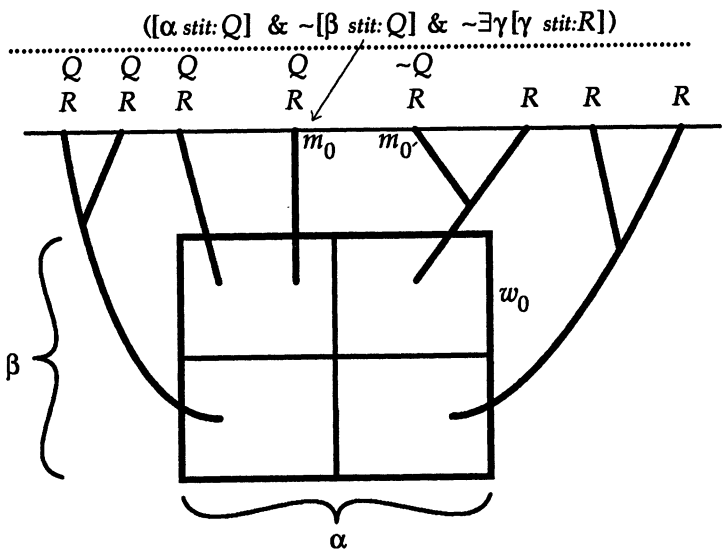


Figure 3: Paradigm *stit*

The picture indicates convenient *given* assignments of values to atomic sentences *below* the dotted line (absence of an indication for an atomic sentence does not mean that it has no value, but only that such value as it may have is irrelevant for the purpose at hand). The values of sentences *above* the dotted line must, however, be *calculated*.

You can see the following from this picture.

- At moment  $m_0$ ,  $\alpha$  sees to it that  $Q$ , that is, the fact that  $Q$  was guaranteed by the prior choice at  $w_0$ . There are two things to look for in support of this calculation.
  1. Positive condition: observe that  $Q$  holds at each of the moments that are choice-equivalent to  $m_0$  at  $w_0$  for  $\alpha$ . (The possible choice for  $\alpha$  at  $w_0$  that is in question is the left half of the box.)
  2. Negative condition: observe that there is a “counter” (as one might say) at  $m_0'$  that testifies that  $w_0$  was a real choice for  $\alpha$

concerning  $Q$ . Had  $\alpha$  made the other choice at  $w_0$ , he or she would have risked the failure of  $Q$  at the instant  $i_{m_0}$ .

- It is easy to see that  $\beta$  did not see to it that  $Q$  at  $m_0$ . (The possible choice for  $\beta$  at  $w_0$  that is in question is the top half of the box.) The failure of  $Q$  at  $m_0'$  shows that the Positive condition is violated, because  $m_0$  and  $m_0'$  are choice-equivalent at  $w_0$  for  $\beta$ .
- It is easy to see that  $w_0$  cannot be used as a witness to anyone's seeing to the fact that  $R$  at  $m_0$ , since at the point  $w_0$ , the truth of  $R$  at the instant  $i_{m_0}$  is already settled. Neither agent  $\alpha$  nor agent  $\beta$  has any control at the moment  $w_0$  over the matter of  $R$  at  $i_{m_0}$ , it having already been settled.
- The two histories emerging from the upper left box signify that what happens after  $w_0$  is partly out of the control of any agent. If you like you may say that what happens is partly "up to Nature," but though it may sometimes make the mathematics easier, I do not think that philosophy is served by passing to a structure in which what is outside of the control of every agent is represented as something that can be "seen to" by Nature.

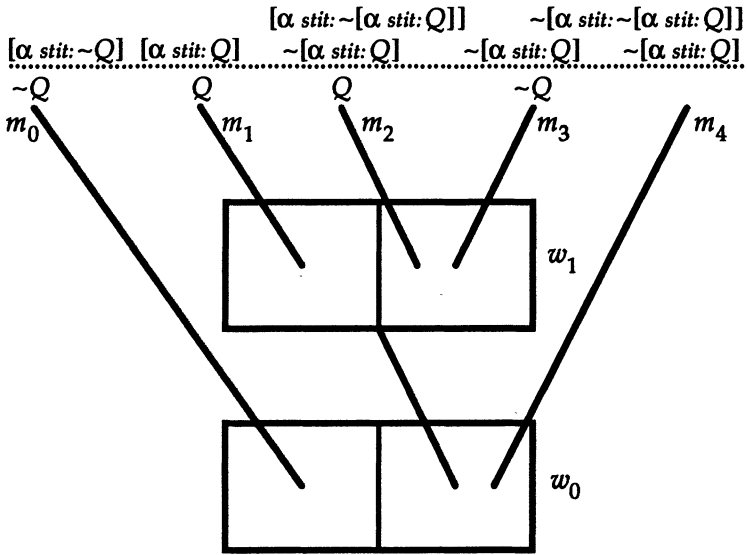
### Failure of *stit* to be closed under logical consequence

Question: is what  $\alpha$  sees to closed under logical consequence? As a paradigm, is it possible that  $\alpha$  sees to the conjunction of  $Q$  and  $R$  but does not see to it that  $R$ ? Of course: as Figure 3 already makes obvious, [ $\alpha$  *stit*:  $Q \& R$ ] holds at the moment  $m_0$ , but [ $\alpha$  *stit*:  $R$ ] does not. The crucial point is that whereas there is a counter  $\sim(Q \& R)$  at  $m_0'$ , there is no counter to  $R$ : as already observed, at the only potential witness  $w_0$ , it is already settled that  $R$  at the relevant instant  $i_{m_0}$ . So what  $\alpha$  sees to is not closed under logical consequence, and obviously so. There is not the slightest paradox in saying, nor any "funny logic" required in calculating, that from the fact that I see to it that an injured man is bandaged it does not follow that I see to it that there is an injured man, even though that an injured man is bandaged logically implies that there is an injured man. To the contrary, it is deeply built into the metaphysics of agency that such cases should be typical.

### Refraining vs. not seeing to it that

Question. Are there any differences between (1) seeing to it that it is false that  $Q$ , (2) the mere absence of seeing to it that  $Q$ , and (3) refraining from seeing to it that  $Q$ , that is seeing to it that you do not see to it that  $Q$ ?

Many foolish things have been said while thinking about this question; Figure 4 vividly illustrates the differences required, and permits the essential calculations. (This and succeeding pictures represent choices only for  $\alpha$ , since only those happen to be relevant to the particular points to be made here.)



**Figure 4:**  $[\alpha \text{ stit}: \sim Q]$  vs.  $\sim[\alpha \text{ stit}: Q]$  vs. refrain ( $[\alpha \text{ stit}: \sim[\alpha \text{ stit}: Q]]$ )

- (1) Easiest to see is that  $[\alpha \text{ stit}: \sim Q]$  holds at  $m_0$ , with witness  $w_0$  and counter at  $m_1$ .
- (2) It will also be useful to note that  $[\alpha \text{ stit}: Q]$  holds at  $m_1$ , with witness  $w_1$  and counter at  $m_3$ . (Pause to observe that  $w_0$  cannot serve as witness for  $[\alpha \text{ stit}: Q]$  at  $m_1$ . Reason:  $Q$  fails at  $m_3$ , which is choice-equivalent for  $\alpha$  to  $m_1$  at  $w_0$ , so that there is a violation of the Positive condition for  $w_0$  to witness  $[\alpha \text{ stit}: Q]$  at  $m_1$ .)
- (3) It is clear that at  $m_2$ , one cannot attribute a guarantee of the fact that  $Q$  holds there to *any* prior choice of  $\alpha$ , for that fact was up to Nature. The same is true of  $m_3$  and  $m_4$ : at all of  $m_2$ ,  $m_3$ , and  $m_4$ ,  $\sim[\alpha \text{ stit}: Q]$  holds, which is the mere absence of seeing to it that. It is worth noticing that we can make this statement about  $m_4$  without even knowing whether or not  $Q$  itself holds there; all we need to observe is the failure of  $Q$  at moment  $m_3$ , which is choice-equivalent to  $m_4$  at  $w_0$  for  $\alpha$ . It follows that the Positive

condition for  $w_0$  to witness  $[\alpha \text{ stit}: Q]$  at  $m_4$  fails in virtue of the failure of  $Q$  at  $m_3$ , and since  $w_0$  is the only potential witness for  $[\alpha \text{ stit}: Q]$  at  $m_4$ , it must be that  $[\alpha \text{ stit}: Q]$  fails at  $m_4$ , which is to say, it must be that  $\sim[\alpha \text{ stit}: Q]$  holds there.

- (4) Moments  $m_2$  and  $m_4$  on the one hand, and  $m_4$  on the other, are quite different with respect to refraining. The picture shows that the moment  $w_1$  *does* stand witness to  $\alpha$ 's responsibility for his or her own inaction with respect to  $Q$  at  $m_2$  or  $m_3$ : not only does the right-hand choice for  $\alpha$  at  $w_1$  guarantee that  $\alpha$  does not see to it that  $Q$ , but the left-hand choice from  $w_1$ , at which  $\alpha$  *does* see to it that  $Q$ , testifies that at  $w_0$   $\alpha$  had a real choice concerning his or her seeing to it that  $Q$ . The moment  $m_1$  stands, that is, as the "counter" required for the truth of the claim that at  $m_2$ , or  $m_3$ ,  $\alpha$  saw to it that he or she did not see to it that  $Q$ .

In contrast to moments  $m_2$  and  $m_3$ , you can tell that in fact at moment  $m_4$ ,  $\alpha$  did *not* actively refrain from not seeing to it that  $Q$ . The only potential witness is  $w_0$ ; but since  $\alpha$  *did* refrain from seeing to it that  $Q$  at  $m_2$ , and since  $m_2$  is choice-equivalent to  $m_4$  at  $w_0$  for  $\alpha$ , the Positive condition fails, and thereby the claim to agency. At  $m_4$  not only does  $\alpha$  fail to see to it that  $Q$ , but he or she also fails to see to it that he or she fails to see to it that  $Q$ . At  $m_4$  you can therefore observe the difference between mere not seeing to something on the one hand, and positively refraining on the other, for, as we have calculated, at  $m_4$  there is *not seeing to it* without *refraining*.

### Could have done otherwise.

It is an obvious feature of our metaphysics that there is an open future, and that if  $[\alpha \text{ stit}: Q]$ , then there is an alternate moment at which  $Q$  is false, namely the counter, and therefore an alternate moment at which  $\sim[\alpha \text{ stit}: Q]$  (since no one can see to the false). This might be and sometimes is expressed in English by saying that "it might have been otherwise," which has nothing to do with agency. But concerning "could have done otherwise" there remains the following question: given an open future, is it true that if  $\alpha$  sees to the fact that  $Q$ , then (1) he or she could have seen to the fact that  $\sim Q$ , or if not that, that at least (2) he or she could have refrained from seeing to it that  $Q$ ? Figure 5, which is just a slight variant of Figure 4, shows that both theses are false, and shows (in a sense I do not know how to define) why those theses are false. The picture permits us to calculate.

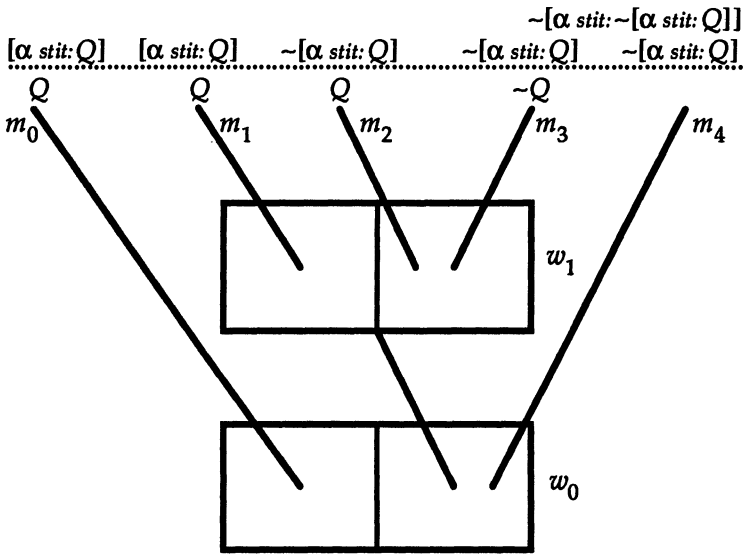


Figure 5: Could have done otherwise

Evidently  $[\alpha \text{ stit}: Q]$  holds at  $m_0$ , with  $w_0$  as witness and  $m_3$  as counter. The rest of the picture is just like Figure 4.

But in the first place, the other choice available to  $\alpha$  at  $w_0$  obviously does not guarantee that  $\sim Q$ , so on that ground alone it was impossible for  $\alpha$  to see to it that  $\sim Q$ . That other possible choice only *risks*  $\sim Q$ , but does not guarantee it.

And in the second place, the other choice available to  $\alpha$  at  $w_0$  does not even guarantee that  $[\alpha \text{ stit}: \sim[\alpha \text{ stit}: Q]]$ , i.e., that  $\alpha$  refrains from seeing to it that  $Q$ . It is possible, but it is not guaranteed, for Nature can take us straight to  $m_4$ , where we calculated that  $\alpha$  does not refrain. If this is what “could have done otherwise” means, then “could have done otherwise” is by no means a consequence of taking the future as open.

One can also see, however, that at  $w_0$  there exists a strategy for  $\alpha$  such that if (a)  $\alpha$  knows about that strategy, and if (b)  $\alpha$  wishes to follow it, and if (c)  $\alpha$  does not run into problems of weakness of the will, then  $\alpha$  is in a position at  $w_0$  in this somewhat Pickwickian or conditional sense of (or absence of sense of) “guarantee” that he or she does not see to it that  $Q$ , that is, that  $\sim[\alpha \text{ stit}: Q]$ . The strategy is simply to make the right-hand choice at each of  $w_0$  and  $w_1$ : then, no matter what Nature has in store, the issue is bound to be  $\sim[\alpha \text{ stit}: Q]$ . But surely you can agree that this weakened state is a long way from what your average expert on free will might have meant by “could have done otherwise,” though perhaps it is what the most subtle dialecticians of the topic were getting at. In any event, it is apparent that the pictures make the discussion easier to follow.

### Closure of *stit* under $\supset$

Question: is what you see to closed under material “implication”? Oddly enough, as Gupta pointed out, it depends on the relative order of the witnesses provided for  $[\alpha \textit{ stit}: P]$  and for  $[\alpha \textit{ stit}: P \supset Q]$ , as is clear from Figures 6 and 7:

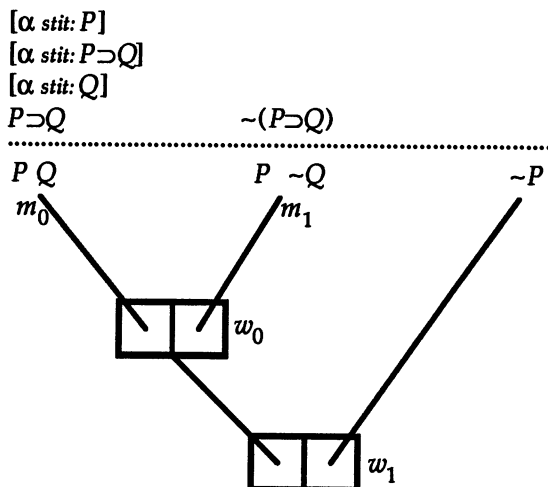


Figure 6: Do  $[\alpha \textit{ stit}: P]$  and  $[\alpha \textit{ stit}: P \supset Q]$  imply  $[\alpha \textit{ stit}: Q]$ ?

In Figure 6 the witness  $w_1$  for the seeing to it that  $P$  at  $m_0$  is earlier than (or the same as) the witness  $w_0$  for the seeing to it of the conditional at  $m_0$ ; in that circumstance one is bound to have  $[\alpha \textit{ stit}: Q]$  at  $m_0$  witnessed by  $w_0$ , and with the same counter at  $m_1$  serving for both  $[\alpha \textit{ stit}: P \supset Q]$  and  $[\alpha \textit{ stit}: Q]$ .

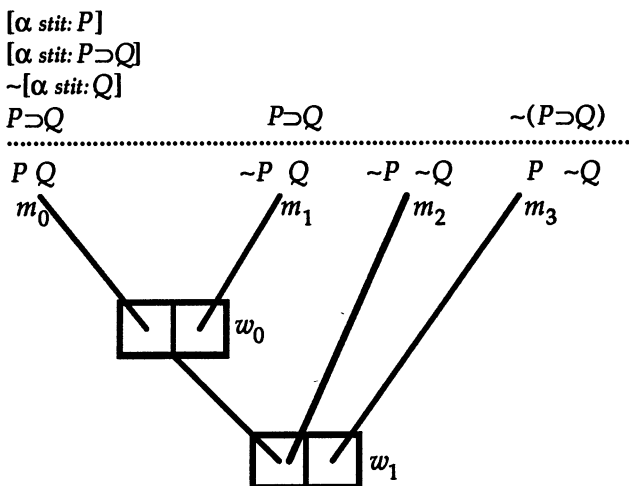


Figure 7: Do  $[\alpha \textit{ stit}: P]$  and  $[\alpha \textit{ stit}: P \supset Q]$  imply  $[\alpha \textit{ stit}: Q]$ ?



In Figure 7, however, the witness  $w_0$  for the seeing to it that  $P$  at  $m_0$  is properly later than the witness  $w_1$  for the seeing to it of the conditional at  $m_0$ . In this case it can be at  $m_0$  that one sees to it that  $P$  and sees to it that  $P \supset Q$  without seeing to it that  $Q$ . In particular,  $w_1$  cannot witness  $[\alpha \text{ stit}: Q]$  at  $m_0$  because of the failure of  $Q$  at  $m_2$ , which is choice-equivalent to  $m_0$  at  $w_1$  for  $\alpha$ . And  $w_0$  cannot witness  $[\alpha \text{ stit}: Q]$  at  $m_0$ , because by then the fact that  $Q$  at  $i_{m_0}$  is settled—there is no “counter.”

### The ten-minute mile

Question: how is it possible to be the agent of one’s own run of a ten-minute mile?

The answer is in the following picture.

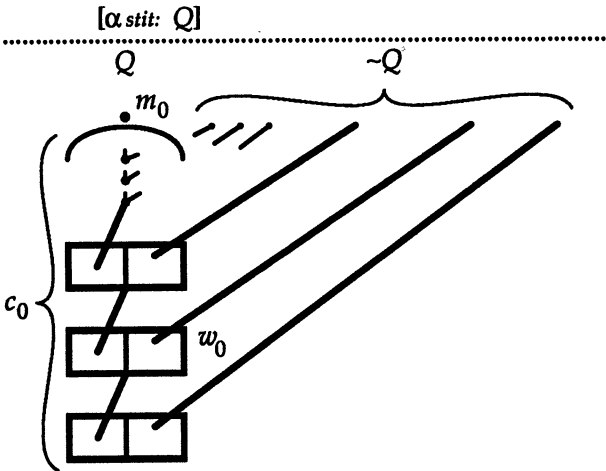


Figure 8: The ten-minute mile

Suppose that  $\alpha$  has been steadily running at a ten-minute pace, and at frequent moments (of which there is no last—this is *the* critical condition)  $\alpha$  has the option to drop out of the run. Consider  $[\alpha \text{ stit}: Q]$  at  $m_0$  as “ $\alpha$  sees to it that  $\alpha$  finishes the mile in just ten minutes.” Evidently  $[\alpha \text{ stit}: Q]$  *should* be true at  $m_0$ , but it is equally evident that *no single prior moment* such as  $w_0$  is adequate as a witness. The reason that  $w_0$  cannot serve as a witness is not just intuitive, though it is that as well. In addition, the picture shows that the Positive condition is violated, for  $Q$  fails at a moment that comes out of a right-hand side of a box that is *later* than  $w_0$ , and hence  $Q$  fails at a moment that is choice-equivalent to  $m_0$  at  $w_0$  for  $\alpha$ .

We therefore need to complicate our semantics (the underlying metaphysics remaining unchanged) by permitting *chains* as well as single moments to count as witnesses. The details are a little delicate, but you can catch the idea. It is the whole chain of choices coming right up to the finish line that stands as witness to the truth at  $m_0$  of “ $\alpha$  sees to it that  $\alpha$  finishes

the mile in just ten minutes.” One has only to generalize the positive and negative conditions appropriately.<sup>19</sup>

The picture shows, incidentally, that that the successful Ten-Minute Miler is, in the sense defined just below, a Busy Chooser!

### Refraining from refraining I

Question: suppose you want to refrain from refraining from seeing to it that you juggle sixteen balls in the air. Is there a way in which you can do that without seeing to it that you juggle sixteen balls in the air? I am reminded of the imperative with which Davidson used to begin each delivery of his intriguing lecture on whether animals can think: he always used to say, “Don’t bother me with your stories!” Now that there is a picture and a calculation, that is what I say: don’t bother me with your stories about locked rooms or making yourself drunk or having your friend or enemy bind you hand and foot, for except in what are impressively complicated circumstances, it is just not possible for you to refrain from refraining from seeing to it that you juggle sixteen balls without your actually seeing to it that you juggle sixteen balls.

Here is the hard fact underlying my Davidsonian impatience with story-tellers: if you can tell a story in which  $[\alpha \text{ stit: } \sim[\alpha \text{ stit: } \sim[\alpha \text{ stit: } Q]]]$  holds but  $[\alpha \text{ stit: } Q]$  does not, then it is going to have to be part of your story that it contains a *Busy Chooser*, that is, an agent  $\alpha$  for whom there is an infinite chain of *nonvacuous* choices occurring in some definite interval (bounded both above and below by moments).<sup>20</sup> But if you allow me to impose on your stories the condition that there are no Busy Choosers, then you cannot tell a story in which  $\alpha$  refrains from refraining from seeing to it that  $\alpha$  juggles sixteen balls but nevertheless fails to see to it that  $\alpha$  juggles sixteen balls. Without Busy Choosers, refraining from refraining from seeing to indeed implies seeing to.

The proof in full is not appropriate to this sketch, but the following picture will give you the flavor. (Since this picture is part of a proof instead of an example, the “dotted line” convention is not wanted.)

---

<sup>19</sup> We are defining what it is for a chain to witness the truth of  $[\alpha \text{ stit: } Q]$  at  $m_0$ . Of course the chain, call it  $c_0$ , must be entirely prior to  $m_0$ . The Negative condition is easy: at every moment  $w_0$  in  $c_0$ ,  $Q$  must fail at some member of  $i_{m_0}$  lying above  $w_0$ . (Observe that this condition does not prevent the choice at  $w_0$  in  $c_0$  from being vacuous, for we do not say that the history on which the counter for  $w_0$  lies must split off at  $w_0$  itself.) For the Positive condition, first say that a moment  $m_1$  in  $i_{m_0}$  is choice-equivalent to  $m_0$  at  $c_0$  for  $\alpha$  if, for every moment  $w_0$  in  $c_0$  that lies below  $m_1$  (hence below both  $m_0$  and  $m_1$ ),  $m_1$  is choice-equivalent to  $m_0$  at  $w_0$  for  $\alpha$ . Then the Positive condition is just that  $Q$  shall hold at every moment that is choice-equivalent to  $m_0$  at  $c_0$  for  $\alpha$ .

<sup>20</sup> That this concept is the relevant one was pointed out by M. McCullagh.

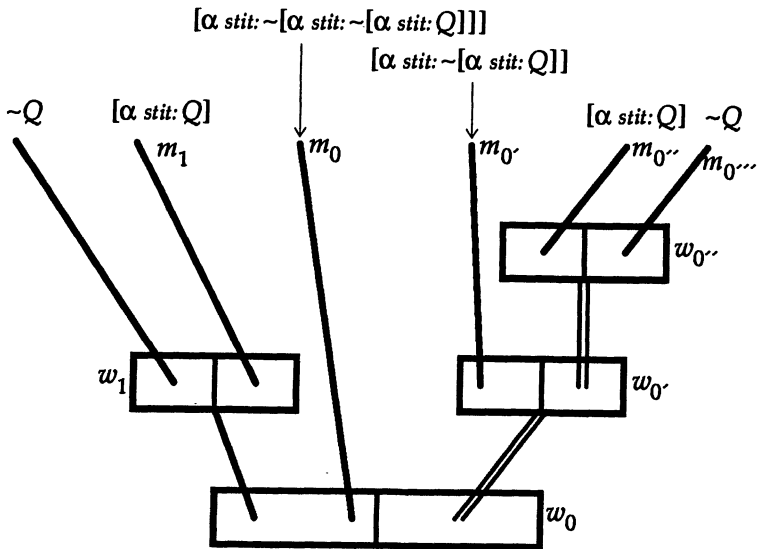


Figure 9: With no Busy Choosers,  $[\alpha \text{ stit: } \sim[\alpha \text{ stit: } \sim[\alpha \text{ stit: } Q]]$  implies  $[\alpha \text{ stit: } Q]$

Suppose that  $[\alpha \text{ stit: } \sim[\alpha \text{ stit: } \sim[\alpha \text{ stit: } Q]]$  holds at  $m_0$ . It needs a witness  $w_0$  and a counter  $[\alpha \text{ stit: } \sim[\alpha \text{ stit: } Q]]$ , which we write at  $m_0$ . This in turn needs a witness  $w_0'$  and a counter  $[\alpha \text{ stit: } Q]$ , which we write at  $m_0'$ ; a definite argument is needed, however, that  $w_0'$  is correctly drawn as later than (or identical to, a possibility expressed by the double lines)  $w_0$ . A similar argument as to the need for (and placement of) a witness and a counter justifies the remainder of the right side of the diagram. Then a *reductio* argument permits us to argue that  $[\alpha \text{ stit: } Q]$  can be “moved over” to a moment  $m_1$  on the left that is choice-equivalent to  $m_0$  at  $w_0$  for  $\alpha$ . The left side of the diagram is part of a subsidiary *reductio*: the picture as drawn places the witness  $w_1$  for  $[\alpha \text{ stit: } Q]$  at  $m_1$  as properly above  $w_0$ , which can be shown to be impossible, provided  $\alpha$  is not a Busy Chooser. In fact, if there are no Busy Choosers, the witness for  $[\alpha \text{ stit: } Q]$  at  $m_1$  must be  $w_0$  itself. This easily implies that  $[\alpha \text{ stit: } Q]$  must be true at  $m_0$ , as desired.

## Refraining from refraining II

Question: what happens to refraining from refraining when one does allow a story to make reference to a Busy Chooser?

Here is a picture that reveals the failure of the implication from  $[\alpha \text{ stit: } \sim[\alpha \text{ stit: } \sim[\alpha \text{ stit: } Q]]$  to  $[\alpha \text{ stit: } Q]$ .<sup>21</sup>

<sup>21</sup> There is not enough room to show all relevant assignments; the idea is that aside from the top right-most moment, those jumping out of the right-hand side of a box have  $Q$ , and those that are limit points of an infinite chain of Busy Choosing have  $\sim Q$ .

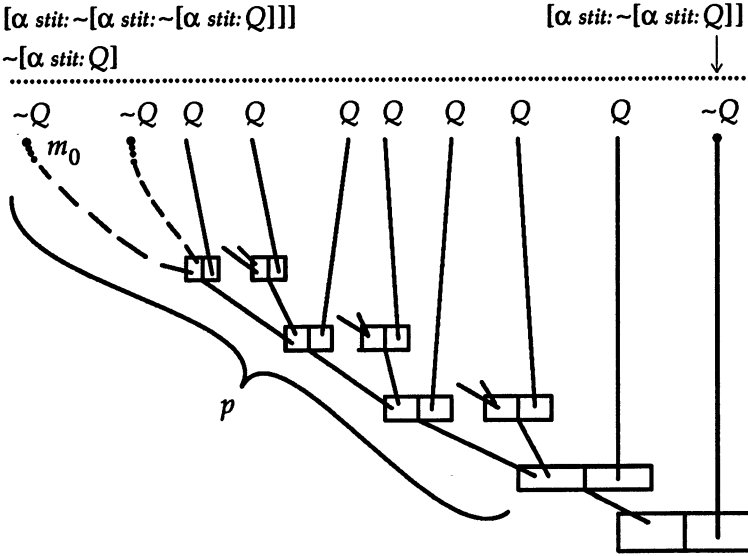


Figure 10: With Busy Choosers,  $[\alpha \text{ stit}: \sim[\alpha \text{ stit}: \sim[\alpha \text{ stit}: Q]]]$  does not imply  $[\alpha \text{ stit}: Q]$

You can tell that it deeply involves a Busy Chooser. I am not going to discuss this picture, partly for lack of space, but partly because I do not understand it myself in intuitive terms. I just want to exhibit what at least one counterexample looks like, so that perhaps someone else can find something simpler than this awesome garden of forking paths.

\* \* \* \* \*

Three points come after the pictures. The first is that in spite of a too sparse history, there is satisfactory internal sense to a modal logic of agency set in a context of a theory of branching time. (I take it to be obvious that since one already has branching time, one can add tense constructions as desired in the way described by Thomason.) The second point is that you ought to think about such a logic before you try to think about how agentives embed in all the philosophically important contexts that require them, for example imperatives, obligation contexts, and plans. The third point is that in fact it is important that we as philosophers take up the half of the theory of agentives neglected by the prevalent too-soon-ontological approach. We should think carefully about how it is that agentives embed in larger linguistic contexts. It is more than a pun to suggest that doing so can help us think about how it is that agents relate to their interpersonal, social, and moral contexts.

## References

- Anderson, A. R., "Logic, norms, and roles." *Ratio*, vol. 4 (1962), pp. 36–49.
- Åqvist, L., *Performatives and verifiability by the use of language*. Philosophical Studies, no. 14, University of Uppsala, Philosophical Society and the Department of Philosophy, Uppsala, 1972.
- \_\_\_\_\_, "A new approach to the logical theory of actions and causality." In: *Logical theory and semantic analysis, Essays dedicated to Stig Kanger on his fiftieth birthday*, ed. S. Stenlund, Dordrecht (D. Reidel Publishing Company), 1974, pp. 73–91.
- \_\_\_\_\_, "An analysis of action sentences based on a 'tree' system of modal tense logic." In: *Papers on tense, aspect and verb classification*, ed. C. Rohrer, Tübingen (TBL Verlag Gunter Narr), 1978, pp. 111–61.
- Åqvist, L. and Mullock, P., *Causing harm: a logico-legal study*. Berlin and New York (Walter de Gruyter), 1989.
- Belnap, N., "Before refraining: concepts for agency." *Erkenntnis*, vol. 34 (1991), pp. 237–69.
- Belnap, N. and Perloff, M., "Seeing to it that: a canonical form for agentives." *Theoria*, vol. 54 (1988), pp. 175–99. W. Edelberg pointed out to us that the informal semantic account of *stit* is garbled in this paper; the account is correct in the version of this paper republished in H. E. Kyburg, Jr., R. P. Loui, and G. N. Carlson (eds.), *Knowledge representation and defeasible reasoning*, Studies in cognitive systems, vol. 5, Dordrecht/Boston/London (Kluwer Academic Publishers), 1990, pp. 167–90; a correct account also occurs in Belnap 1989.
- Bennett, J., *Events and their names*. Indianapolis (Hackett Publishing Company), 1988.
- Bressan, A., *A general interpreted modal calculus*. New Haven (Yale University Press), 1972.
- Castañeda, H.-N., *La lógica general de las normas y la ética*. Universidad de San Carlos, Guatemala, vol. 30, 1954.
- \_\_\_\_\_, "The paradoxes of deontic logic: the simplest solution to all of them in one fell swoop," 1981. In: Hilpinen 1981, pp. 37–86.
- Chellas, B. F., *The logical form of imperatives*. Stanford, CA (Perry Lane Press), 1969.
- Chisholm, R., "Symposium: action, motive, and desire. The descriptive element in the concept of action." *The journal of philosophy*, vol. 90 (1964), pp. 613–25.
- \_\_\_\_\_, "Human freedom and the self." The Lindley Lecture, Department of Philosophy, University of Kansas, 1964a, pp. 3–15. Reprinted in K. Lehrer (ed.), *Freedom and determinism*, New York (Random House),

- 1966, pp. 11–44; and in R. Chisholm, *On metaphysics*, Minneapolis (University of Minnesota Press), 1989, pp. 5–15.
- Danto, A., *Analytical philosophy of action*. Cambridge (Cambridge University Press), 1973.
- Davidson, D., “The logical form of action sentences.” In: *The logic of decision and action*, ed. N. Rescher, Pittsburgh (Pittsburgh University Press), 1967, pp. 81–95.
- \_\_\_\_\_, *Essays on actions and events*. London and New York (Oxford University Press), 1980.
- Dazeley, H. and Gombocz, W., “Interpreting Anselm as logician.” *Synthese*, vol. 40 (1979), pp. 71–96.
- Elgesem, D., “Intentions, actions and routines: a problem in Krister Segerberg’s theory of action.” Unpublished manuscript, 1989.
- Fitch, F. B., “A logical analysis of some value concepts.” *The journal of symbolic logic*, vol. 28 (1963), pp. 135–42.
- Goldman, A., *A theory of human action*. Princeton (Princeton University Press), 1970.
- Gupta, A., *The logic of common nouns: an investigation in quantified modal logic*. New Haven (Yale University Press), 1980.
- Hamblin, C. L., *Imperatives*. Oxford and New York (Basil Blackwell), 1987.
- Henry, D., “Saint Anselm on the varieties of ‘doing’.” *Theoria*, vol. 19 (1953), pp. 178–83.
- \_\_\_\_\_, “Saint Anselm’s *De ‘Grammatico’*.” *Philosophical quarterly*, vol. 10 (1960), pp. 115–26.
- \_\_\_\_\_, *The logic of Saint Anselm*. Oxford (Oxford University Press), 1967. See especially §4, which presents among other things Anselm’s proposal (in a work composed *circa* 1100) that “the analysis of ‘doing’ will in fact be an analysis of  $x$ ’s doing so that  $p$  ... .”
- Hilpinen, R., “On the semantics of personal directives.” *Ajatus*, vol. 35 (1973), pp. 140–57.
- \_\_\_\_\_ (ed.), *New studies in deontic logic: norms, actions, and the foundations of ethics*. Dordrecht (D. Reidel Publishing Company), 1981.
- Hohfeld, W. N., *Fundamental legal conceptions as applied in judicial reasoning, and other essays*. New Haven (Yale University Press), 1919.
- Humberstone, I. L., “An alternative account of bringing about.” Unpublished manuscript, 1976. Read to a meeting of the Australasian Association for Logic in 1976.
- \_\_\_\_\_, “An alternative account of bringing about (with a pinch of relevance).” *Bulletin of the Section of Logic*, vol. 6 (1977), pp. 144–45. Reprinted in *The Relevance Logic Newsletter*, vol. 2 (1977), pp. 107–8. This is an abstract of Humberstone 1976.

- Kanger, S., *New foundations for ethical theory*. Stockholm, 1957. Reprinted in R. Hilpinen (ed.) *Deontic Logic: Introductory and Systematic Readings*, Dordrecht (D. Reidel Publishing Company) 1971, pp. 36–58.
- \_\_\_\_\_, “Law and logic.” *Theoria*, vol. 38 (1972), pp. 105–32.
- Kanger, S. and Kanger, H., “Rights and parliamentarism.” *Theoria*, vol. 6 (1966), pp. 85–115.
- Kenny, A., *Action, emotion and will*. London (Routledge and Kegan Paul), 1963.
- Lindahl, L., *Position and change: a study in law and logic*. Dordrecht (D. Reidel Publishing Company), 1977.
- Makinson, D., “On the formal representation of rights relations.” *Journal of philosophical logic*, vol. 15 (1986), pp. 403–25.
- McCall, S., “Objective time flow.” *Philosophy of science*, vol. 43 (1976), pp. 337–62.
- Mullock, P., “Causing harm: criminal law.” *Law and philosophy*, vol. 7 (1988), pp. 67–105.
- Needham, P., *A semantic approach to causal logic*. Master’s thesis, Birmingham University, 1971.
- Pörn, I., *The logic of power*. Oxford (Basil Blackwell), 1970.
- \_\_\_\_\_, *Elements of social analysis*. Philosophical Studies No. 10, Uppsala (Philosophical Society and the Department of Philosophy, University of Uppsala), 1971.
- \_\_\_\_\_, “Some basic concepts of action.” In: *Logical theory and semantic analysis, Essays dedicated to Stig Kanger on his fiftieth birthday*, ed. S. Stendlund, Dordrecht (D. Reidel Publishing Company), 1974, pp. 93–101.
- \_\_\_\_\_, *Action theory and social science: some formal models*. Dordrecht (D. Reidel Publishing Company), 1977.
- Pratt, V. R., “Application of modal logic to programming.” *Studia logica*, vol. 39 (1980), pp. 257–74.
- Prior, A., *Past, present, and future*. Oxford (Oxford University Press), 1967.
- Segerberg, K., “Applying modal logic.” *Studia logica*, vol. 39 (1980), pp. 275–95.
- \_\_\_\_\_, “Action-games.” *Acta philosophica fennica*, vol. 32 (1981), pp. 220–31.
- \_\_\_\_\_, “The logic of deliberate action.” *Journal of philosophical logic*, vol. 11 (1982), pp. 233–54.
- \_\_\_\_\_, “Towards an exact philosophy of action.” *Topoi*, vol. 3 (1984), pp. 75–84.
- \_\_\_\_\_, “Routines.” *Synthese*, vol. 65 (1985), pp. 185–210.

- \_\_\_\_\_, "Models for actions." In: *Analytical philosophy in comparative perspective*, ed. B. K. Matilal and J. L. Shaw, Dordrecht (D. Reidel Publishing Company), 1985a, pp. 161–71.
- \_\_\_\_\_, "Action, logic and social theory." *Acta philosophica fennica*, vol. 38 (1987), pp. 282–98.
- \_\_\_\_\_, *Validity and satisfaction in imperative logic*. Auckland philosophy papers, University of Auckland, Department of Philosophy, 1988.
- \_\_\_\_\_, "Talking about actions." *Studia logica*, vol. 47 (1988a), pp. 347–52.
- \_\_\_\_\_, *Getting started: beginnings in the logic of action*. Auckland Philosophy Papers, University of Auckland, Department of Philosophy, January, 1989.
- Talja, J., "A technical note on Lars Lindahl's *Position and change*." *Journal of philosophical logic*, vol. 9 (1980), pp. 167–83.
- Thomason, R. H., "Indeterminist time and truth-value gaps." *Theoria*, vol. 36 (1970), pp. 264–81.
- Thomson, J., *Acts and other events*. Ithaca (Cornell University Press), 1977.
- von Kutschera, F., "Grundbegriffe der Handlungslogik." In: *Handlungstheorien interdisziplinär* vol. I, *Handlungslogik, formale und sprachwissenschaftliche Handlungstheorien*, ed. H. Lenk, Munich (Wilhelm Fink Verlag), 1980, pp. 67–106.
- \_\_\_\_\_, "Bewirken." *Erkenntnis*, vol. 24 (1986), pp. 253–81.
- von Neumann, J. and Morgenstern, O., *Theory of games and economic behavior*. Princeton (Princeton University Press), 1944.
- von Wright, G. H., *Norm and action. A logical inquiry*. London (Routledge and Kegan Paul), 1963.
- \_\_\_\_\_, "On the logic of norms and actions," 1981. In: Hilpinen 1981, pp. 3–35.
- Walton, D., "Modal logic and agency." *Logique et analyse*, vol. 69–70 (1975), pp. 103–11.
- \_\_\_\_\_, "St. Anselm and the logical syntax of agency." *Franciscan studies*, vol. 36 (1976), pp. 298–312.
- \_\_\_\_\_, "Logical form and agency." *Philosophical studies*, vol. 29 (1976a), pp. 75–89.
- \_\_\_\_\_, "On the logical form of some commonplace action expressions." *Grazer philosophische studien*, vol. 10 (1980), pp. 141–48.