BRIAN F. CHELLAS

Time and Modality in the Logic of Agency*

Abstract. Recent theories of agency (*sees to it that*) of Nuel Belnap and Michael Perloff are examined, particularly in the context of an early proposal of the author.

1. Introduction

More than twenty years ago, in *The logical form of imperatives*,[1] I advanced an account of agency expressed in terms of an operator meant to correspond to the expression *sees to it that*. Since then, a number of authors have fashioned further theories of agency along similar lines, and recently, in a series of papers, Nuel Belnap and Michael Perloff have put forth a new account. Their theories of agency are complex, fascinating, and illuminating — without a doubt the most subtle and sophisticated proposals of their kind to date.

*Elements of this paper formed the contents of lectures I gave in New Zealand and Australia in 1989. I would like again to thank Graham Oddie at Massey University, in Palmerston North, Jack Copeland at the University of Canterbury, in Christchurch, John Bacon at the University of Sydney, and Graham Priest at the University of Queensland for their kindness and the gracious receptions they and their colleagues gave me. My thanks go to Graham Oddie and Krister Segerberg, who organized a workshop on "Events, Processes, Actions" at Lake Taupo, New Zealand, in November 1989, and invited me to participate with a preliminary version of this paper. In June 1990, I presented a fuller specimen at the annual meeting of the Society for Exact Philosophy, in Wakulla Springs, Florida. Bob Beard at Florida State University organized that splendid gathering, and I am grateful to him for the opportunity to speak at it.

For the past several years, Nuel Belnap has sent me copies and updates of his and Michael Perloff's papers. I would like to record my gratitude to him for this and also, especially, for extended comments on the penultimate draft of the present paper. With a few exceptions, I have not tried to take these into account here; I hope that discussions of points on which we disagree will find their way into print in due course.

I would also like to acknowledge and thank a referee for a number of helpful suggestions.

My largest debt is to Krister Segerberg, who as professor and head of the philosophy department at the University of Auckland invited me to spend a sabbatical autumn (antipodal spring) with him in 1989. It was he who suggested — and then insisted — that I contribute to his seminar on modal logic and agency a session or two on Belnap and Perloff's theories, and then he encouraged me to write this paper. I would like to express my deep gratitude as well to Krister and Anita Segerberg for their hospitality and companionship during my stay in New Zealand.

[1] [6], referred to hereafter as *LFI*.

Studia Logica **51**: 485–517, 1992.

My purpose in this paper is to set out and examine these theories.[2]

Following this introduction, the paper continues with a section on preliminary matters, succeeded by one setting forth some possible principles in the logic of agency. The next two sections are an exposition of the account in *LFI.* Then come five sections on Belnap and Perloff's original theory, followed by three sections that examine this critically. The penultimate section presents a revised theory, and the paper ends with some concluding remarks.

In presenting these theories and my own, I have done what I can to minimize technicality, in the hope that this will enhance intelligibility and make the subject more accessible to readers less formally inclined. I have not avoided formality altogether, since the subject admits of precise and unambiguous formulations.

2. Preliminaries

In *LFI* the operator Δ is introduced to create formulas $\Delta\tau\Phi$, where τ represents a singular term and Φ a formula. In this exposition I shall write $\Delta\alpha A$ instead and at the same time, except for an example later on, restrict attention to nonquantificational idioms. Informally, and with the usual confusion, $\Delta\alpha A$ is to be read "α sees to it that A".

Belnap and Perloff use the operator *stit* for *sees to it that*, and write "*stit* sentences" as $[\alpha\, stit{:}Q]$. For the sake of simplicity and uniformity I shall write $STIT\alpha A$ instead.

Belnap and Perloff promulgate several theses about $STIT$ sentences, three of which bear mentioning here. The first is the *Stit Complement Thesis*, according to which:

> $STIT\alpha A$ is grammatical and meaningful (though perhaps silly) for an *arbitrary* sentence A.

The second, dubbed the *Agentiveness of Stit Thesis*, says:

> $STIT\alpha A$ is always agentive in α, regardless of its complement.

[2]The papers I have in mind are [3] and [4], by Belnap and Perloff, and Belnap's [1] and [2]. (I do not consider "The way of the agent", Belnap and Perloff's contribution to the present volume.) In [2] Belnap thoroughly revises the theory along lines adumbrated in [1] and hinted at in [3] and [4]. For my purposes here it is best to begin with Belnap and Perloff's original ideas, and then consider the later theory. In setting out the theories I have omitted a great deal, though I believe without misrepresenting the authors' views. Because only [3] and [4] have appeared in print, I quote section titles in lieu of page numbers when giving references to [1], [2], [3], and [4].

The third is the *Stit Paraphrase Thesis*:

> A is agentive in α just in case A is paraphrasable as (or is strongly equivalent to) $STIT\alpha A$.

We shall return to these theses below.[3]

3. Some principles of agency

It will be useful in what follows to keep in mind some possible principles of the logic of agency, in terms of which one can appreciate differences among competing, sometimes incompatible theories of agency. The list that follows is minimal, and yet it has enough items to provide a healthy basis for such appreciation. So as to be neutral between Δ and *STIT*, and to cover some other modalities as well, let us use $\Box$ to represent *sees to it that.*

RE. $\dfrac{A \leftrightarrow B}{\Box\alpha A \leftrightarrow \Box\alpha B}$

M. $\Box\alpha(A \wedge B) \rightarrow (\Box\alpha A \wedge \Box\alpha B)$

C. $(\Box\alpha A \wedge \Box\alpha B) \rightarrow \Box\alpha(A \wedge B)$

N. $\Box\alpha\top$

$\overline{\text{N}}$. $\neg\Box\alpha\top$

T. $\Box\alpha A \rightarrow A$

4. $\Box\alpha A \rightarrow \Box\alpha\Box\alpha A$

Q. $\Box\alpha\Box\beta A \rightarrow \Box\alpha A$

The reader will doubtless notice the similarity of several of these principles to ones familiar in modal logic. Nevertheless, brief commentary is called for.

The first principle, the rule of inference RE, posits a principle of extensionality. Intuitively, it says that if A and B are equivalent then so are the results $\Box\alpha A$ and $\Box\alpha B$ of "modalizing" them. Schemas M and C express the inward and outward distributivity of agency operators, and N (in the presence of RE) stipulates that every agent sees to anything that holds logically or trivially.

With RE present the schema M supplies the rule of inference

$$\text{RM.} \quad \frac{A \rightarrow B}{\Box\alpha A \rightarrow \Box\alpha B}.$$

[3]See [1], "Who wants a logic of *stit* (sees to it that)?" Though all *STIT* sentences are agentive, the category goes beyond these; compare [3] and [4], "Introduction" and "Agentives" in both.

According to this, agency is closed under simple consequence: if B is a consequence of A, then $\Box\alpha B$ is a consequence of $\Box\alpha A$.[4]

Schema C adds to RE and M to produce a stronger rule of closure under consequence :

$$\text{RR.} \quad \frac{(A \wedge B) \rightarrow C}{(\Box\alpha A \wedge \Box\alpha B) \rightarrow \Box\alpha C}.$$

Thus if a sentence C is a consequence of any finite, nonempty collection of sentences, then $\Box\alpha C$ likewise follows from the collection of their $\Box\alpha$–modalizations.

Finally, with N in addition to RE, M and C, a general rule of consequence emerges :

$$\text{RK.} \quad \frac{(A_1 \wedge \cdots \wedge A_n) \rightarrow A}{(\Box\alpha A_1 \wedge \cdots \wedge \Box\alpha A_n) \rightarrow \Box\alpha A}.$$

When $n = 0$, RK is just

$$\text{RN.} \quad \frac{A}{\Box\alpha A}$$

— a rule of "necessitation".

Taken together, the four principles RE, M, C, and N mean that the logic of the operator is *normal*.[5]

The sentence $\overline{\text{N}}$, on the other hand, alleges the opposite of N, to wit (again in the context of RE) that no agent ever sees to something that is necessarily so. N and $\overline{\text{N}}$ are at the bottom of a controversy in the logic of agency, to which we shall return.

The schema T is fundamental: an agent sees to it that something is the case only if it is the case. By truth–functional reasoning, T yields

$$\text{O.} \quad \neg(\Box\alpha A \wedge \Box\beta\neg A)$$

— according to which it is impossible that one agent should see to it that something is so while another sees to it that it is not so. With the help of RE and M, O in turn delivers a sentence that states that no agent can see to the impossible :

$$\text{P.} \quad \neg\Box\alpha\bot.$$

Schema 4 is less controversial than, e.g., N or $\overline{\text{N}}$, but is included here as it represents another point of difference between Belnap and Perloff's account and that of *LFI*.

[4]Notice that by itself RM yields both RE and M.

[5]RK alone yields a normal modal logic. See e.g. [8], pp. 111 – 130 and 231 – 245, for more on principles and characterizations of normality.

Schema Q, finally, expresses a kind of principle of responsibility: if agent α sees to it that agent β sees to it that something is the case, then *ipso facto* α sees to it (too).[6]

4. Semantics for Δ

Models for the language provide interpretations and truth values in terms of *histories, times, agents*, and certain *relations* among these.

For our purposes it is enough that *time* be represented by any linear ordered set, so that we can speak unambiguously of one time being earlier than or later than another. We construe *histories* as functions from times to events, or states of affairs, which are themselves otherwise unspecified. So $h(t)$, or h_t, is the state of affairs in history h at time t.[7]

Two important relations between histories can be distinguished, for each time t. The first relates histories having the *same past* at t, and we write :

$$h =_t h' \text{ iff } h_{t'} = h'_{t'}, \text{ at every time } t' < t.$$

This relation appears in *LFI.* The other relation, which is implicit in Belnap and Perloff's account, connects histories having the *same present and past* at t :

$$h \equiv_t h' \text{ iff } h_{t'} = h'_{t'}, \text{ at every time } t' \leq t.$$

The relations $=_t$ and $\equiv_t$ are useful in the depiction of histories branching as they move into the future, from earlier to later times; for though we differ on certain details, I agree with Belnap and Perloff that at any point of time a history's past is unique, whereas its future may be manifold. Excluded is the idea that histories different in the past ever come together in the future, since that would make for more than one past at the histories' temporal meeting point. In *LFI* I nevertheless did not confine the set of histories to those with pasts that are unique at all times; i.e. I did not impose a *future branching only* condition :

$$\text{if } h_t = h'_t \text{ then } h \equiv_t h'.$$

[6] Q encapsulates the legal maxim *Qui facit per alium facit per se* (my thanks to Noyes Leech for this). We say "(too)" because by T if α sees to it that β sees to it then β sees to it — so that both α and β see to it.

[7] "Event" is what I said in *LFI*; see p. 81. I wish I had said "state of affairs", which is what I did say in [7], p.121, and what I will say here. In *LFI* the set of times was taken to be the set of integers, $\{\ldots, -1, 0, +1, \ldots\}$ (p.70). We shell write $t < t'$ to mean that time t is earlier than time t', and $t \leq t'$ to mean that t is no later than t'.

In order to forestall consideration of histories with divergent pasts, I instead relied on "historical relevance" conditions on relations used to interpret individual modalities, such as the agency operator Δ.[8]

In terms of this picture, it is helpful to think of the *future cone* determined by a state of affairs — viz. the collection of historical paths branching into the future from the state of affairs. Thus if histories h and h' are identical up to and including a time t, i.e. $h \equiv_t h'$, then h and h' belong to the future cone emanating from h_t (equally, from h'_t).

Another element of a model relates histories with respect to agents and times. For histories h and h', $R^{\alpha}_t(h, h')$ means that for agent α, at time t, h' is an *instigative alternative* to h, i.e. that h' is "under the control of — or responsive to the actions of" α at t. The relation R^{α}_t is defined for each agent α and time t. It is reflexive and is subject to a *historical relevance* condition, viz. that instigative alternatives for an agent at a time must have the same past at that time :

$$R^{\alpha}_t(h, h') \text{ only if } h =_t h'.$$

In other words, instigative alternatives to h for α at t must all lie within the future cone emanating from h_t. In *LFI* it is this condition that is fundamental to the conception of agency as dependent on the past, for it means that the states of affairs so to speak under an agent's control are limited to those that are possible outcomes of the way the world has been up to (but not including) a moment of agency.[9]

Lastly, a *valuation* in a model assigns agents to agent symbols and truth values to sentences relative to history–time pairs (h, t).[10]

We write $\models^h_t$ for *is true at* (h, t). Evaluations for truth–functional constructions are straightforward and familiar. Truth conditions for the operator Δ are expressed :

$$\models^h_t \Delta\alpha A \text{ iff } \models^{h'}_t A \text{ for every history } h' \text{ such that } R^{\alpha}_t(h, h').$$

That is to say, $\Delta\alpha A$ — "α sees to it that A" — is true in a history at a time just in the case A is true at all instigative alternatives to the history for the agent at the time.

[8] Same–present–and–past relations appear explicitly e.g. in [16]. If histories are always *future branching only* (as in Belnap and Perloff's account), the definition of these relations would be simply: $h \equiv_t h'$ iff $h_t = h'_t$.

[9] See *LFI*, pp.63, 78 – 79, and 82. The relation there is R^*.

[10] We are following Belnap and Perloff's practice in using the same symbols for agents and agent symbols. For their conception of agents, see [1], "Semantics for *stit*"; [2], "Theory of agents and choices"; and [3] and [4], "Semantics".

The better to understand the logic of Δ in what follows, let us introduce a "historical necessity" operator $\boxdot$ to go with the relations $=_t$:

$$\models_t^h \boxdot A \text{ iff } \models_t^{h'} A \text{ for every history } h' \text{ such that } h =_t h'.$$

Thus $\boxdot A$ is reckoned true in a history at a time just when A is true in all histories having the same past at the time.

In anticipation of developments below, we may recognize a companion to $\boxdot$, another "historical necessity" operator $\circledcirc$, to go along with the relations $\equiv_t$:

$$\models_t^h \circledcirc A \text{ iff } \models_t^{h'} A \text{ for every history } h' \text{ such that } h \equiv_t h'.$$

So $\circledcirc A$ holds at (h,t) just in case A holds at all pairs (h',t) for which h and h' are identical at t and all times prior to t.

Later on we shall encounter as well some operators for tense and other temporal constructions.[11]

Finally, A is *valid* just when it is true at all history–time pairs throughout all models; A *implies* B means that $A \to B$ is valid; and A and B are *equivalent* just in case each implies the other.

5. Logic of Δ

What valid and validity–preserving principles emerge from the semantics for Δ in *LFI*?

It can be seen that Δ is a normal modal operator, i.e. that the following principles hold for Δ:

RE. $$\frac{A \leftrightarrow B}{\Delta\alpha A \leftrightarrow \Delta\alpha B}$$

M. $\Delta\alpha(A \wedge B) \to (\Delta\alpha A \wedge \Delta\alpha B)$

C. $(\Delta\alpha A \wedge \Delta\alpha B) \to \Delta\alpha(A \wedge B)$

N. $\Delta\alpha\top$

Because of the reflexivity of R_t^α, moreover,

T. $\Delta\alpha A \to A$

[11] $\boxdot$ appears in *LFI*; $\circledcirc$ does not. See [16] for an analogue of $\circledcirc$.

is valid, and the schema

$$\text{Q.} \quad \Delta\alpha\Delta\beta A \rightarrow \Delta\alpha A$$

also holds (e.g. apply RM to T). Indeed the only items on the list that fail for Δ are $\overline{\mathsf{N}}$ $(\neg\Delta\alpha\top)$, which contradicts N, and 4 $(\Delta\alpha A \rightarrow \Delta\alpha\Delta\alpha A)$.

To the extent that English counterparts of RE, M, C, T, and Q have the ring of validity, the account in *LFI* is virtuous to uphold these principles. Validity is denied to 4, but it must be acknowledged that English renderings of this strike the ear at least ambivalently. Validity of 4 would mean that the relations R_t^α are transitive. I cannot think why that should be so, although it may be that an analysis of instigative alternativeness would yield this.

Both the historical necessity operators $\boxdot$ and $\boxcircle$ are also normal, and, as each of the relations $=_t$ and $\equiv_t$ is an equivalence, $\boxdot$ and $\boxcircle$ also obey the laws of the modal logic S5. Because of the condition the historical alternatives include instigative ones, the following schema is valid :

$$\text{N}^*. \quad \boxdot A \rightarrow \Delta\alpha A.$$

Note that N for Δ follows from this together with $\boxdot\top$ (i.e. N for $\boxdot$). So it may be said that in this modeling every agent sees to whatever must be so, logically by N, as well as historically by N*.[12]

These results, that agents are to be held accountable for what must be the case historically, let alone logically, are difficult to believe. Belnap and Perloff find N and N* objectionable, as now I do. Further on we shall see that we nonetheless part ways on how to avoid the likes of these.

Apart from the presence of N and N*, the account of agency embodied in Δ would seem to be minimal. Yet some of the remaining principles are the subject of dispute, as we shall see.

Let me turn now to theories of agency of Belnap and Perloff.

6. A theory of *STIT*

Belnap and Perloff fashion an account of *STIT* within a "metaphysical backdrop" of branching time and starting from the idea that $STIT\alpha A$ means

[12] See *LFI*, pp. 65 – 67 and 85 – 87. Regarding transitivity and the validity of 4, and for more information about S5, see e.g. [8], pp. 76 – 82 and 138 – 140. Because histories having the same past and present have the same past, i.e. because $\equiv_t$ is included in $=_t$, $\boxdot A \rightarrow \boxcircle A$ is also valid.

that the truth of A is somehow "guaranteed by a prior choice of α".[13]

At bottom, their "metaphysics" differs from that in *LFI*. Superficially, or at least some distance from the bottom, we both conceive of histories as branching toward the future in time. But I take time to be a single linear progression, whereas they regard time itself as branching. For them *moments* are the fundamental temporal units, and under a relation of *earlier than* they form a future–branching tree. Histories, being paths through moments, are the tree's branches, and a time, an *instant* of time in their parlance, is a maximal collection of moments none of which is earlier or later than any of the others.

Ways of comparing Belnap and Perloff's metaphysics with mine will suggest themselves. For example, the states of affairs in *LFI* correspond closely to moments. However, in *LFI* one and the same state of affairs can occur at different times and in different histories, while moments are all distinct. Another difference arises from the fact that in *LFI* each history and time determines a state of affairs, but for Belnap and Perloff, though each (instant of) time has at least one moment in some history, in principle there can exist histories and times that determine no moment at all. In practice, however, they avoid this possibility by stipulating that all histories (i.e. moment–seqences) be isomorphic.[14]

There is a further, semantic difference to be noted. Whereas I evaluate sentences in terms of histories and times, Belnap and Perloff do so in terms of histories and moments. But given that, for Belnap and Perloff, moments are always associated with histories and (instants of) times, there should be no problem about using histories and times as evaluation points instead.[15]

This is the course I follow here, employing the models of *LFI* and continuing to use history–time pairs as points of evaluation, to formulate Belnap and Perloff's truth conditions for *STIT* sentences. So far as I can see, these revisions occasion no pertinent alteration of their interpretation of *STIT* and no alteration in the set of *STIT* sentences valid in their account. This way of proceeding makes it easier to compare the theories of Belnap and Perloff with others, that of *LFI* in particular, and reveals how theirs in fact constitute less of a departure than they think from some earlier theories of the

---

[13][1], "Semantics for *stit*". [4], "Semantics", has "guaranteed by", where [3] has "fully due to".

[14]See e.g. [2], "Theory of instants". In all of [1], [2], [3], and [4], the authors cite [13] and [15] as influences on their thinking about branching time and histories.

[15]Belnap writes, "I wholly follow Thomason [15] in believing that a semantics in the context of historical indeterminism needs to make truth...relative to moment/history pairs", and "the moment alone does not determine a truth value" ([2], "Witness of *stit* by moments"). Compare [3] and [4], "Semantics".

modal logic of agency.

As mentioned, Belnap and Perloff's interpretation of *STIT* sentences is founded on a notion of *choice*, for an agent at a moment. An agent's choice is a collection of histories stemming from a single history at some moment, or "choice point". Thus choices for an agent at a time are always within that moment's future cone. Such choices moreover do not overlap, and collectively they exhaust the histories in the moment's future cone. In short, an agent's choices at a choice point partition the future cone for that moment. Where an agent has but one choice at choice point, i.e. where the choice is identical with the future cone, the agent is said to have a vacuous, or Hobson's choice.

We can represent this notion of a choice for an agent at a time by means of equivalence relations between histories. Thus for agent α and time t, $E_t^\alpha(h, h')$ if and only if histories h and h' are together in a choice for α at t.

Belnap and Perloff put three constraints on choice equivalence relations. The first, indicated above, is a *historical relevance* constraint to the effect that choices fall within future cones :

$$E_t^\alpha(h, h') \text{ only if } h \equiv_t h'.$$

The second constraint figures importantly in Belnap and Perloff's theory. Belnap calls it the *no choice between undivided histories* condition, and it says that once a history is part of a choice it remains so for ever. The effect of this is to keep subsequent branchings of a history in a choice from straying outside the choice. We can state the condition thus :

$$\text{if } t' < t, \text{ then } h \equiv_t h' \text{ only if } E_{t'}^\alpha(h, h').$$

Another way of putting this is to say that a future cone subsequent to a choice point is always inside a single choice determined at that point.[16]

The third constraint is called *something happens* and states that at any choice point

> for every way of selecting one possible choice for each agent from among his or her set of choices, the intersection of all possible choices selected must contain at least one history.

As Belnap and Perloff note, this condition is relevant only where there is more than one agent. It has at least one controversial implication, as we shall see below in connection with the schema Q.[17]

---

[16] See [1], "Semantics for *stit*", and [2], "Theory of agents and choices".

[17] The quotation is from [2], "Theory of agents and choices"; the wording in [1] ("Semantics for *stit*") differs slightly. Compare [3] and [4], "Semantics", where the condition is called *the-world-goes-on*.

The notion that *seeing to it* essentially involves an idea of prior choice is spelled out in Belnap and Perloff's account of what it means for $STIT\alpha A$ to be true at a given moment, or history–time point, (h,t). To wit: there is a prior choice point (h,t') such that (1) A is true at t throughout the choice for α at (h,t') that contains h (i.e. in every history choice equivalent to h at t'), and (2) A is false at t somewhere in the choice point's future cone (i.e. in some historical alternative to h stemming from (h,t')). Thus, intuitively, α's seeing to it that A on a given occasion is determined by a choice for α on a prior occasion — a choice that is non–trivial in the sense that at the time of choice A was not bound to be true anyway.[18]

Here are the truth conditions for $STIT$ sentences stated formally :

$\models_t^h STIT\alpha A$ iff there is a time $t' < t$ such that
(1) $\models_t^{h'} A$ for every history h' such that $E_{t'}^{\alpha}(h,h')$
and
(2) $\not\models_t^{h'} A$ for some history h' such that $h \equiv_{t'} h'$.

Belnap refers to clause (1) as the *positive condition*, and to (2) as the *negative condition*. Notice that although Belnap and Perloff allow for the possibility of a Hobson's, or vacuous, choice, where an agent's choice is the same as the future cone, by the *negative condition* they do not admit agency based on such a choice. So $STIT$ may be better read "sees to it — *really*".[19]

7. Extensionality and settledness

In formulating truth conditions for agency sentences using Δ and $STIT$, we have stuck to the idea of truth at history–time pairs (doing duty in Belnap and Perloff's case for history–moment pairs). But this idea may fall short of capturing all that might be meant by truth at the states of affairs or moments associated with such pairs.

Let us say that a sentence A satisfies the *extensionality* condition just in the case its truth value at any history–time pair (h,t) depends solely on the associated state of affairs h_t ; in other words, just in case :

$$\text{if } h_t = h'_t, \text{ then } \models_t^h A \text{ if and only if } \models_t^{h'} A.$$

Notice that if A satisfies the *extensionality* condition then it also satisfies a *weak extensionality* condition :

$$\text{if } h \equiv_t h', \text{ then } \models_t^h A \text{ if and only if } \models_t^{h'} A.$$

[18] See [1], "Semantics for *stit*", and [2], "Witness of *stit* by moments"; compare [3] and [4], "Semantics".

[19] Again, see [1], "Semantics for *stit*", and [2], "Witness of *stit* by moments", and compare [3] and [4], "Semantics".

If *future branching only* holds, as it does for Belnap, the two conditions are equivalent.

Should valuations be extensional in the sense that atomic sentences are made to satisfy the *extensionality* condition? This would yield the result that all sentences of purely truth–functional composition, as well as some others, also satisfy the condition.[20]

Even if atomic sentences are assumed to satisfy *extensionality*, it cannot be expected that every sentence will. In general, sentences that contain interfering modalities, such as certain future tense constructions, will not. If this is regarded as a shortcoming of a theory that assigns truth values relative to history–time pairs, then what may be wanted is an account of a sentence's being "settled true" (in Belnap's phrase) at a state of affairs or moment — i.e. true not just at (h,t) but at every (h',t) such that $h \equiv_t h'$.[21]

Let us write $\Vert\!\models_t^h A$ to mean that A is (not only true but) *settled true* at (h,t). The definition is :

$$\Vert\!\models_t^h A \text{ iff } \models_t^{h'} A \text{ for every history } h' \text{ such that } h \equiv_t h'.$$

The contrary, *settled falsity*, may be defined :

$${}_t^h\!=\!\!\mid\!\mid A \text{ iff } \not\models_t^{h'} A \text{ for every history } h' \text{ such that } h \equiv_t h'.$$

We say that a sentence is simply *settled* at (h,t) just in the case it is settled true or settled false at the pair.

It is apparent that sentences satisfying the *extensionality* condition are settled at all history–time pairs.

In the definiens of his truth conditions for $STIT$ sentences, Belnap employs settledness, not merely truth and falsity.[22] Thus :

$\models_t^h STIT\alpha A$ there is a time $t' < t$ such that
(S1) $\Vert\!\models_t^{h'} A$ for every history h' such that $E_{t'}^{\alpha}(h,h')$
and
(S2) $\Vert\!\not\models_t^{h'} A$ for every history h' such that $h \equiv_{t'} h'$.

But mere truth is sufficient: where a time t' is earlier than a time t, the *positive conditions* (1) and (S1) in the definientia hold equally, and so do

[20] The *extensionality* condition appears in *LFI* (p. 92, n. 3) and is also found in [7] (p. 123), [15] (in effect, p.277), and in [16] (see pp. 138 – 139 and 146). Belnap and Perloff presumably assume it, although this is not wholly clear.

[21] Or such that $h_t = h'_t$ in Belnap and Perloff's scheme, since their histories do not branch toward the past.

[22] At least in [2], "Witness of *stit* by moments". Earlier statements of truth conditions, e.g. in [1], "Semantics for *stit*", and [4], "Semantics", are ambiguous.

the *negative conditions* (2) and (S2). In one direction this is obvious, since truth follows from settled truth. For at least the *positive conditions* the other direction is worth proving, as the argument provides the first evidence of the importance of the *no choice between undivided histories* condition.

Assume that t' is earlier than t and that (1) holds. For *positive condition* (S1), suppose that $E^{\alpha}_{t'}(h, h')$, and to show then that A is settled true at (h', t), suppose that $h' \equiv_t h''$ and argue that A is true at (h'', t). By *no choice between undivided histories*, the suppositions that $t' < t$ and $h' \equiv_t h''$ imply that $E^{\alpha}_{t'}(h', h'')$. So by transitivity of choice equivalence, $E^{\alpha}_{t'}(h, h'')$. From this together with *positive condition* (1) it follows that A is true at (h'', t).

We omit the argument for the interesting direction between *negative conditions* (2) and (S2).

It is clear therefore that we may continue to use truth and falsity rather than settledness in evaluating *STIT* sentences. However, we shall see that the presence or absence of the *extensionality* condition is critical in connection with the equivalence of certain forms of sentences about agency.

The reader will doubtless have noticed that the ideas of settled truth and falsity are definable using the same–past–and–present historical necessity operator. For at a history–time pair (h, t) a sentence A is settled true if and only if $\boxdot A$ is true, and A is settled false just in case $\boxdot \neg A$ is true. So in leaving the topic let us note that the equivalence of truth and settled truth in the truth conditions for $STIT\alpha A$ means that the operator $\boxdot$ can appear without effect against the content A. That is, $STIT\alpha A$ is equivalent to $STIT\alpha \boxdot A$.[23]

8. Witnesses

Belnap refers to certain choice moments as "witnesses".[24] Witnesses for a *STIT* sentence true at an evaluation point (h, t) may be identified as the history–time pairs (h, t') associated with the times t' earlier than t that satisfy the *positive* and *negative* clauses of the truth conditions. A question

[23]Likewise, $\Delta\alpha A$ is equivalent to $\Delta\alpha \boxdot A$, though this does not rest on a *no choice between undivided histories* condition. Note that any sentence A satisfying the *extensionality* condition is equivalent to $\boxdot A$. See [16], pp. 136 – 146.

[24]In a similar vein he refers to a moment associated with (h', t) in the *negative condition*, where A is false, as a "counter" — in [1], "*Stit* pictures", where the "witness" terminology emerges. Belnap and Perloff speak of a "choice–point" in [3], which becomes a "witnessing 'choice–point' " in [4] ("Semantics" in both).

then arises: How many witnesses can there be for a $STIT$ sentence true at a given evaluation point?

The answer is just one: witnesses are unique. This emerges as a corollary to the following *witness identity lemma*:

> If A implies B and $\models_t^h STIT\alpha A$ and $\models_t^h STIT\beta B$, then no witness for $STIT\alpha A$ can be earlier than any witness for $STIT\beta B$.

To see this, suppose that A implies B and $STIT\alpha A$ and $STIT\beta B$ both hold at (h,t). Suppose further, for reductio, that $STIT\alpha A$ has a witness (h,t_0) earlier than a witness (h,t_1) for $STIT\beta B$. Then by the *no choice between undivided histories* postulate all the histories deriving from (h,t_1) are within the choice for α at (h,t_0) that contains h. But A holds at t throughout the histories in this choice, and hence so does B, since A implies B. If B is true throughout all the histories possible from (h,t_1) onward, however, then no choice for any agent at (h,t_1) can be genuine, since there is no historical alternative outside the choice in which B is false at t. Thus $STIT\beta B$ is false in h at t, contrary to our assumption.

As a corollary to the *witness identity lemma*, we have :

> If A is equivalent to B and $\models_t^h STIT\alpha A$ and $\models_t^h STIT\beta B$, then the witnesses for $STIT\alpha A$ and $STIT\beta B$ are the same.

For if A and B are equivalent, then by the lemma no witness for $STIT\alpha A$ can be earlier or later than any for $STIT\beta B$. So the witnesses are the same.

As a corollary to this, taking $A = B$, we obtain uniqueness :

> If $\models_t^h STIT\alpha A$, then there is exactly one witness for $STIT\alpha A$.

The witness identity results underscore how it is that, for Belnap and Perloff, present agency depends upon the past history of an agent. But uniqueness of witnesses may occasion disquiet, for it means that a choice cannot be subsequently overturned, that even the agent is powerless to choose differently later on. Perhaps looking at choices as *options* for an agent reduces uneasiness felt in the face of their immutability.[25]

[25]Belnap speaks of choices "open for" an agent ([2], "Theory of agents and choices") and sometimes refers to choices as options (e.g. in [1], "Semantics for *stit*").

9. *STIT* defined

Reflection on the truth conditions shows that *STIT* sentences have a structure something like $\langle \mathsf{P}\rangle(\triangle\alpha A \wedge \neg \boxdot A)$, where $\langle \mathsf{P}\rangle$ is a past tense operator meaning "it (at least) once was the case that", the operator $\triangle$ corresponds to the choice equivalence relations E^α_t, and $\boxdot$ is the same–past–and–present historical necessity operator. Truth conditions for $\langle \mathsf{P}\rangle$ and $\triangle$ are as follows :

$$\models^h_t \langle \mathsf{P}\rangle A \quad \text{iff there is a time } t' < t \text{ such that } \models^h_{t'} A.$$

$$\models^h_t \triangle\alpha A \quad \text{iff } \models^{h'}_t A \text{ for every history } h' \text{ such that } E^\alpha_t(h,h').$$

However suggestive, $\langle \mathsf{P}\rangle(\triangle\alpha A \wedge \neg \boxdot A)$ does not do the job of $STIT\alpha A$, since truth conditions for this sentence read :

$\models^h_t \langle \mathsf{P}\rangle(\triangle\alpha A \wedge \neg \boxdot A)$ iff there is a time $t' < t$ such that
(1) $\models^{h'}_{t'} A$ for every history h' such that $E^\alpha_{t'}(h,h')$
and
(2) $\not\models^{h'}_{t'} A$ for some history h' such that $h \equiv_{t'} h'$.

The difficulty is that in the definiens of the truth conditions for $STIT\alpha A$ at a point (h,t) both E^α and $\equiv$ are indexed by a time t' earlier than t, whereas A itself is evaluated at t itself. By contrast, in the truth conditions for $\langle \mathsf{P}\rangle(\triangle\alpha A \wedge \neg \boxdot A)$, the sentence A is evaluated in terms of the earlier time t' rather than t. In other words, in the truth conditions for the proposed structure there is a temporal shift throughout the definiens rather than in just parts of it.

The problem admits of resolution if *temporal double indexing* of the truth predicate is introduced, so that truth conditions are given in terms of two temporal indices instead of one. Where h is a history and t and n are times, we write $\models^h_{t,n}$ for *is true at* (h,t) *with respect to* n, and we evaluate the operators $\langle \mathsf{P}\rangle$, $\triangle$, and $\boxdot$ as follows :

$$\models^h_{t,n} \langle \mathsf{P}\rangle A \quad \text{iff there is a time } t' < t \text{ such that } \models^h_{t',n} A.$$

$$\models^h_{t,n} \triangle\alpha A \quad \text{iff } \models^{h'}_{t,n} A \text{ for every history } h' \text{ such that } E^\alpha_t(h,h').$$

$$\models^h_{t,n} \boxdot A \quad \text{iff } \models^{h'}_{t,n} A \text{ for every history } h' \text{ such that } h \equiv_t h'.$$

The temporal index n, meant to suggest the present time, or "now", is idle in the truth conditions for these operators (and, of course, in those for truth–functional operators). But where times in a context of evaluation are the same — e.g. in $\models^h_{t,t}$ — the extra temporal index permits persistent reference to the original time t in the evaluation. The usefulness of this will be apparent shortly.

To formulate a precise definition of the *STIT* operator, we need to employ an operator $\textcircled{N}$ — meaning "now" or "at present time" — with truth conditions :

$$\models^h_{t,n} \textcircled{N}A \text{ iff } \models^h_{n,n} A.$$

Here the extra temporal index is not idle, and this solves the problem of "temporal shift" that we encountered with the structure $\langle\mathsf{P}\rangle(\triangle\alpha A \wedge \neg \boxdot A)$. For in an initial context $\models^h_{t,t}$ the effect of $\textcircled{N}$ will be to restore the time t' in a subsequent context $\models^h_{t',t}$ to the original time t.

The desired definition of *STIT* is then :

$$STIT\alpha A = \langle\mathsf{P}\rangle(\triangle\alpha \textcircled{N}A \wedge \neg \boxdot \textcircled{N}A).$$

The following sequence of steps establishes the correctness of this definition :

$\models^h_{t,t} STIT\alpha A$ iff there is a time $t' < t$ such that

(1) $\models^h_{t',t} \triangle\alpha \textcircled{N}A$
and
(2) $\models^h_{t',t} \neg \boxdot \textcircled{N}A$

— i.e. such that

(1) $\models^{h'}_{t',t} \textcircled{N}A$ for every history h' such that $E^{\alpha}_{t'}(h,h')$
and
(2) $\not\models^{h'}_{t',t} \textcircled{N}A$ for some history h' such that $h \equiv_{t'} h'$

— i.e. such that

(1) $\models^{h'}_{t,t} A$ for every history h' such that $E^{\alpha}_{t'}(h,h')$
and
(2) $\not\models^{h'}_{t,t} A$ for some history h' such that $h \equiv_{t'} h'$.

It is readily seen that these last two clauses, apart from the duplication of temporal indices, are identical to those in Belnap and Perloff's truth conditions.[26]

In a review of the history of the logic of agency, Belnap locates *LFI* and a number other works in a tradition of "binary relational semantics". He

[26] See [5], pp. 121 - 124, for more about "two-dimensional" tense logic. Burgess cites [9], [10],[12], and [17] by way of further references.

comments on the failure, in his opinion, of semantics of this sort, saying that "it has *remained* obscure what one is to make of the binary relations that serve as the founding elements of the enterprise", and he judges that "[t]he proper conclusion is rather that one should doubt the likelihood that the semantics themselves can serve in the way that was hoped".[27]

The definition above of $STIT$ shows, however, that this operator is analyzable in terms of "binary relations", very much within the tradition. This is not to gainsay the value of Belnap and Perloff's approach, of course, but rather to put their theory in a perspective more familiar, and perhaps more congenial, to toilers in agentival fields.

10. Logic of *STIT*

What then is the logic of agency as expressed by $STIT$ sentences? From the very statement of truth conditions for $STIT$ sentences, $STIT$ evidently obeys the rule of replacement of logical equivalents :

RE. $$\frac{A \leftrightarrow B}{STIT\alpha A \leftrightarrow STIT\alpha B}.$$

It is also apparent, essentially because of the reflexivity of the choice equivalence relations E_t^α, that the schema

T. $$STIT\alpha A \rightarrow A$$

holds for $STIT$. As remarked above, from T both O and P, i.e. $\neg(STIT\alpha A \wedge \neg STIT\beta A)$ and $\neg STIT\alpha\bot$, follow.

It should also be apparent that

$\overline{\mathsf{N}}$. $$\neg STIT\alpha\top$$

holds, and therefore N ($STIT\alpha\top$) fails, in the logic of $STIT$. This is because of the *negative condition*, the requirement that the content of a true $STIT$ sentence be false at some historical alternative emanating from the witness. To see this clearly, suppose, for reductio, that $STIT\alpha\top$ is true at some history–time pair (h, t). Then the *positive condition* is certainly satisfied, since $\top$ holds everywhere, but the *negative condition* cannot be satisfied, since this would mean that $\top$ is false somewhere. We shall observe further the profound effect the *negative condition* has on the logic of $STIT$.

[27] [1], "Mini–history".

Less evident are the successes for $STIT$ of these schemas :

C. $(STIT\alpha A \wedge STIT\alpha B) \rightarrow STIT\alpha(A \wedge B)$

4. $STIT\alpha A \rightarrow STIT\alpha STIT\alpha A$

For C: Suppose that $STIT\alpha A$ and $STIT\alpha B$ both hold at a history–time pair (h,t). Then A holds there and at t in every other choice equivalent (for α), all of which stem from a witness (h,t'), where t' is some time prior to t. Likewise, B is true at (h,t) and at t in every other choice equivalent (for α) that stems from a witness (h,t''), where t'' is also prior to t. If the witnesses (h,t') and (h,t'') are the same, i.e. if $t' = t''$, then relevant choice equivalents are the same, and the conjunction $A \wedge B$ likewise holds at t in this histories. Our supposition implies too the presence of a counter for $STIT\alpha B$, a pair (h',t) within the witness's future cone, where B is false. This pair also rejects the conjunction $A \wedge B$ and so provides a counter for $STIT\alpha(A \wedge B)$. Thus $STIT\alpha(A \wedge B)$ holds at (h,t). Suppose, however, that witnesses (h,t') and (h,t'') are not the same — let us say t' is earlier than t''. Then by *no choice between undivided histories* the future cone and hence the choice equivalents (for α) emerging from the later witness (h,t'') all lie within α's choice at the earlier witness (h,t'). This means that both A and B are true at t throughout the choice equivalents stemming from the witness (h,t''). In the same way as before, this witness's future cone contains a counter for $STIT\alpha B$ that also counters $STIT\alpha(A \wedge B)$. Therefore, again, $STIT\alpha(A \wedge B)$ holds at (h,t).

For 4: Assume that $STIT\alpha A$ it true at (h,t), to show that $STIT\alpha STIT\alpha A$ also holds at this pair. Our assumption means that $STIT\alpha A$ is witnessed by (h,t'), where t' is earlier than t, and this witness leads both to (h,t) and the other choice equivalents (for α) and to a counter (h',t). The sentence A is true at t in the choice equivalents and false at the counter. It follows from this that (h,t') witnesses $STIT\alpha A$ at t in all the choice equivalents, since (h',t) is equally a counter in each case. Noting this, and that $STIT\alpha A$ is false at the counter (because A is), we conclude that $STIT\alpha STIT\alpha A$ also holds at t in all the choice equivalents, in particular at (h,t).

Belnap and Perloff do not mention the validity of C or 4, though the presence of C among the theses for $STIT$ surely counts in favor of the theory. Presumably they would welcome 4, since in conjunction with T it yields the validity of

$$STIT\alpha A \leftrightarrow STIT\alpha STIT\alpha A.$$

That is to say, the success of 4 provides confirmation of a special case of the *Stit Paraphrase Thesis*.[28]

Let us consider next the *Qui facit per alium* principle :

Q. $STIT\alpha STIT\beta A \rightarrow STIT\alpha A$

FOR Q: If the agents α and β are identical then the schema is just a special case of T. So let us suppose from now on that α and β are different. The argument is complicated, but worth setting out in view of the importance of Q and because the reasoning gives a good idea of the subtlety and ingenuity of Belnap and Perloff's ideas — and incidentally shows again the significance of the *no choice between undivided histories* constraint (through its role in the *witness identity* results).[29]

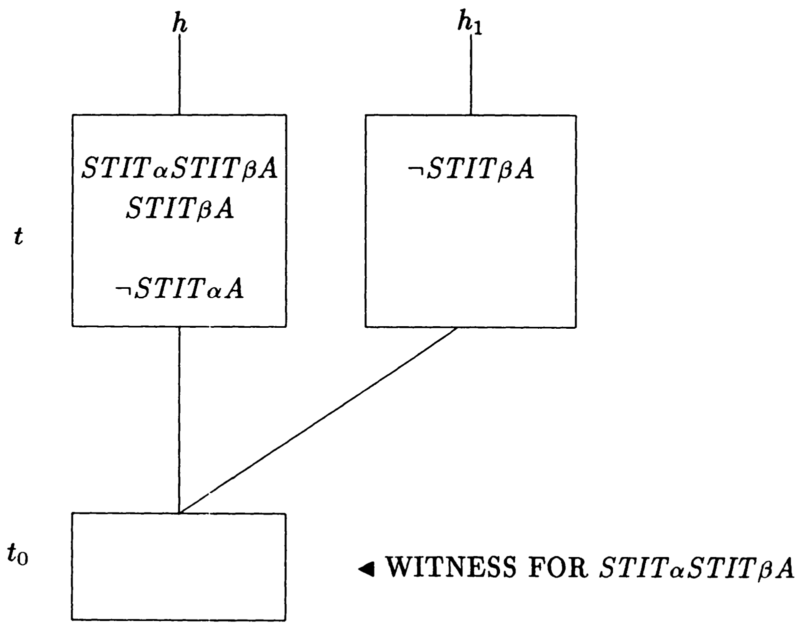

Fig. 1

[28]"A is agentive in α just in case A is paraphrasable as (or is strongly equivalent to) $STIT\alpha A$" (section 2, above; recall that all $STIT$ sentences are agentive).

[29]Because (as we shall see) $STIT$ does not obey the rule RM, the simple argument of applying RM to T is unavailable.

We proceed by reductio. Assume that Q is false at a pair (h,t). Then $STIT\alpha STIT\beta A$ holds at (h,t) while $STIT\alpha A$ fails. This means that there is a witness (h,t_0) for $STIT\alpha STIT\beta A$ (t_0 earlier than t), that at t $STIT\beta A$ holds in h and all other choice equivalents for α, and that there is counter (h_1,t) within the witness's future cone where $STIT\beta A$ is false. A picture is valuable at this juncture; see figure 1.

Because $STIT\beta A$ holds at (h,t) it must have a witness (h,t'), where t' is earlier than t. There are three places for t': later than t_0; earlier than t_0; the same as t_0.

The first of these — t' later than t_0 — is ruled out by the *witness identity lemma*: Since $STIT\alpha STIT\beta A$ implies $STIT\beta A$ (by T), no witness for the former can precede a witness for the latter.

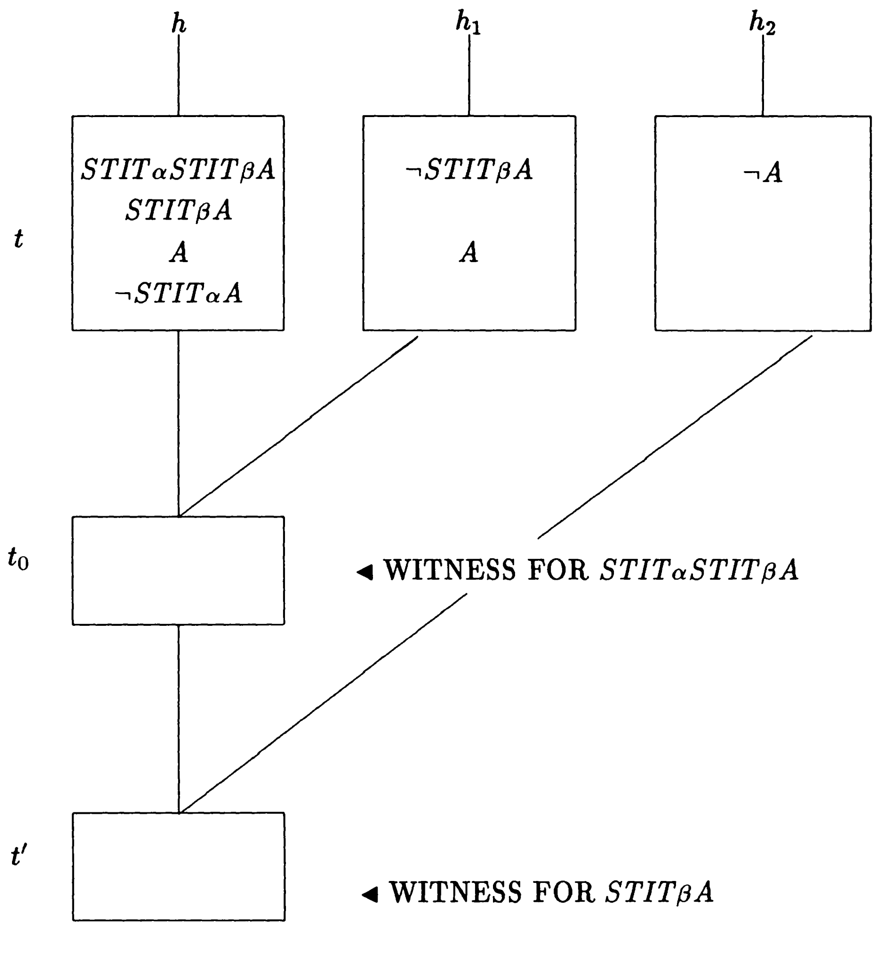

Fig. 2

The second place — t' earlier than t_0 — is also ruled out. For suppose otherwise, i.e. that (h,t') in figure 2 is a witness for $STIT\beta A$ at (h,t). Here we find A true at t in all historical continuations from (h,t_0), since by the *no choice between undivided histories* condition these are all choice equivalents for β from (h,t'). (The pair (h_2,t) is brought in as a counter.) But this means that (h,t') is also a witness (for β) to the truth of $STIT\beta A$ at (h_1,t) — which contradicts its falsity there (because it is a counter for $STIT\alpha STIT\beta A$ at (h,t)).

So we are left with possibility that $t' = t_0$. Then the picture is as in figure 3. The important things to observe here are: first, because $STIT\alpha STIT\beta A$ implies $STIT\beta A$, which implies A, A is true throughout all α's choice equivalents at t with respect to the witness (h,t_0); secondly, at t there is now a historical alternative h_2 to h, deriving from (h,t_0), in which A is false.

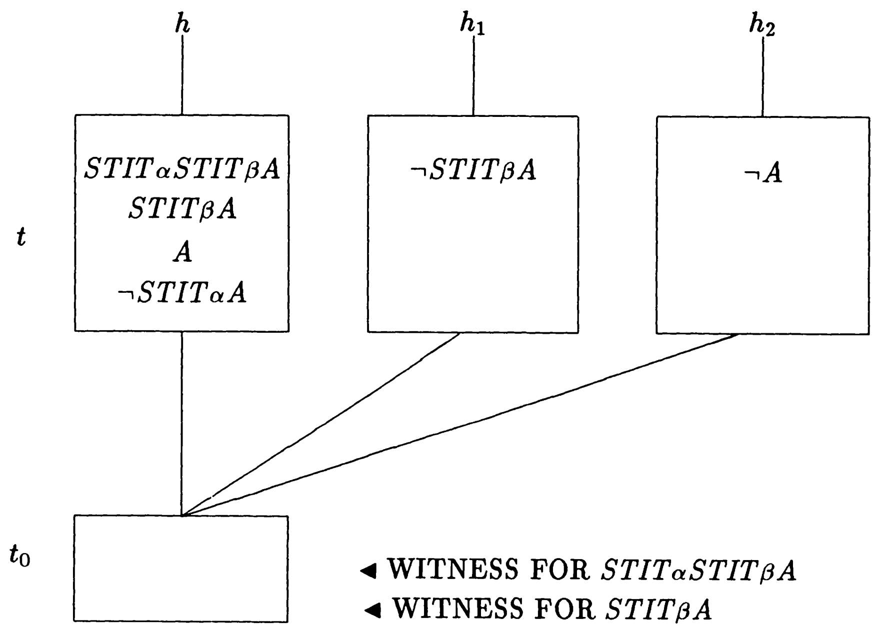

Fig. 3

Let us consider, finally, $STIT\alpha A$, which is supposed to fail at (h,t). This means that at time t, *either* A is false in some choice equivalent for α to h with respect to the witness (h,t_0) — which is ruled out — *or* A is true in all historical alternatives to h emanating from (h,t_0) — to which h_2 provides a counterexample.

Thus none of three placements of a witness for $STIT\beta A$ is tenable, and we conclude therefore that no counterexample exists to the validity of **Q**.

The foregoing argument relies, metaphysically, only on *historical relevance* and *no choice between undivided histories*. So it may come as a disappointment to learn that within Belnap and Perloff's full theory the validity of Q for *STIT* is trivial. For, given the *something happens* condition, Q's antecedent, $STIT\alpha STIT\beta A$, is false whenever the agents α and β are different.[30]

A proof of this proceeds exactly as above for Q up to third and last possibility for the placement of t', viz. that $t' = t_0$ (see figure 3 again). Then we show that given *something happens* this too leads to a contradiction. For where α and β are distinct, the condition tell us that at the witness (h, t_0) each choice for β shares some history with each choice for α. Recall that $STIT\beta A$ holds at t throughout the choice for α that includes h. So at (h, t_0) each of β's choices contains some history in which $STIT\beta A$ is true at t. By our reasoning in the proof for 4, then, $STIT\beta A$ holds at t in every history in all of β's choices at the witness. But this is impossible, since it means that at t $STIT\beta A$ is both true and false at the counter (h_1, t) for $STIT\alpha STIT\beta A$.[31]

It seems to bizarre to deny that an agent should be able to see to it that another agent sees to something. To the extent that the *something happens* condition is responsible for this effect, it should be reexamined and perhaps jettisoned.[32]

Let us conclude this survey of the logic of *STIT* with a few further negative results. We have already seen that N is invalid, but its loss is hardly a matter for mourning. More significant is the invalidity of the principle M and hence (in the presence of RE) the failure of even the weakest rule of closure under consequence :

$$\text{RM.} \qquad \frac{A \to B}{STIT\alpha A \to STIT\alpha B}.$$

One way of seeing that M and RM fail, a way that illuminates the role of $\overline{\mathsf{N}}$, is to note that an application of RM to the tautology $A \to \top$ yields $STIT\alpha A \to STIT\alpha\top$, and given $\overline{\mathsf{N}}$ this would mean that every sentence

[30]The news was disappointing to me, at least. Belnap pointed it out and provided a proof in a letter to me (26 October 1990). See his formulation of the *something happens* condition in section 6 above.

[31]Equally, since (by T) A is true at t throughout β's choices and hence there is no room in the witness's future cone for a history in which A is false at t, i.e. for a counter to $STIT\beta A$ as required by the *negative condition.*

[32]Belnap calls *something happens* a "fierce constraint" ([2], "Theory of agents and choices").

$STIT\alpha A$ was a contradiction.[33]

11. Closure under consequence

The presence of $\overline{\mathsf{N}}$ thus prevents success for M and RM, although it should be emphasized that there are counterexamples for M and RM that involve only contingent contents for their *STIT* sentences. Are there independent reasons for rejecting M and RM?

Belnap addresses this question by means of a modification of the so-called Good Samaritan paradox familiar from deontic logic. He writes 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.[34]

And he alludes to this again, writing :

> There is not 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.[35]

It is clear that formalization of Belnap's example requires at least the apparatus of first order quantification. One way of treating "an injured man" in a simple sentential context uses an existential quantifier. Thus "An injured man is bandaged" comes out as $\exists x(Fx \wedge Gx)$, with F for "injured man" and G for "bandaged", and on one reading "I see to it that an injured man is bandaged" takes the form $STIT\alpha\, \exists x(Fx \wedge Gx)$.

Then using RM one may reason as follows from "I see to it that an injured man is bandaged" to "I see to it that there is an injured man":

1.	$STIT\alpha\, \exists x(Fx \wedge Gx)$	premiss
2.	$\exists x(Fx \wedge Gx) \rightarrow \exists xFx$	logic
3.	$STIT\alpha\, \exists x(Fx \wedge Gx) \rightarrow STIT\alpha\, \exists xFx$	2, RM
4.	$STIT\alpha\, \exists xFx$	1,3, logic

But it should be obvious that the English premiss — "I see to it that an injured man is bandaged" — is misrepresented by the formulation in the

[33] It is worth noting that a weak version of M nevertheless holds for *STIT* : $STIT\alpha(A \wedge B) \rightarrow (STIT\alpha A \vee STIT\alpha B)$.

[34] [1], "Who wants a logic of *stit* (sees to it that)?".

[35] [1], "*Stit* pictures".

first line: the indefinite description "an injured man" is wrongly given narrow scope. Within the resources of *STIT* and first order quantificational formalism, "I see to it that an injured man is bandaged" should be rendered $\exists x(Fx \wedge STIT\alpha Gx)$ — from which there is indeed no inference at all to $STIT\alpha \exists x Fx$.

This rebuttal has even more force when one puts "the" for "an" in the premiss of the example. No one is tempted to assign the resulting *definite* description, "the injured man", narrow scope, precisely because when one does so the (revised) premiss, "I see to it that there is just one injured man and he is bandaged", *does* imply "I see to it that there is an injured man".

Apart from questionable examples like this, 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.

12. Trivial pursuits

We have already noticed the strong effect of the presence of $\overline{\mathsf{N}}$ in Belnap and Perloff's logic of agency, particularly in connection with the presence of RM. Rejection of $\overline{\mathsf{N}}$ and acceptance of RM do not together entail the validity of N. But as noted above, it does mean that $STIT\alpha A \rightarrow STIT\alpha\top$ is valid, and this means that N is true at every history–time pair where any *STIT* sentence is true for any α. The question for some, resolved negatively by Belnap and Perloff, is whether $STIT\alpha\top$ can be true.

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. Done things

What are we to make of sentences such as "John sees to it (now) that yesterday Joan took out the trash" and "Betty sees to it (now) that Boyd once brought his umbrella to work" — constructions in which past tense sentences are filled in following *sees to it that*?

Most of us do not believe that someone can bring about or see to something's being the case in the past in the way that appears to be expressed by sentences like these. "*STIT–past*" (or "Δ*–past*") sentences like "John sees to it (now) that yesterday Joan took out the trash" seem to be false, or anyway true only to the extent that they mean the same as corresponding past or "*past–STIT*" forms such as "Yesterday Joan took out the trash" or "Yesterday John saw to it that Joan took out the trash".

One response would be to rule such sentences out of bounds as ungrammatical or not genuinely agentive. But I wholly agree with Belnap and Perloff on the points embodied in their *Stit Complement* and *Agentiveness of Stit Theses*.[36] Any sentence can complement *sees to it that*; $\Delta\alpha A$ and $STIT\alpha A$ are well formed, meaningful, and agentive for any sentence A. So this response is not available.

In *LFI* there is a happy resolution for a large class of *Δ–past* sentences. Where Ⓟ is any past tense operator like $\langle \mathsf{P} \rangle$, and A satisfies the *extensionality* condition, it turns out that:

$$\Delta\alpha \text{ Ⓟ} A \text{ is equivalent to Ⓟ} A.$$

The reason for this is essentially the *historical relevance* condition in *LFI*: instigative alternatives for an agent at a time all have the same past history at that time, and so a sentence true in the past for one such alternative will be true in the past for all.[37]

At least one class of *STIT–past* sentences is also unproblematic on Belnap and Perloff's account, for the sentences all turn out to be equivalent to their *past–STIT* counterparts. Suppose that time has the structure of the integers, so that each point in time has an immediate predecessor. Then let the meaning of the operator $\ominus$ be given by this truth condition :

$$\models_t^h \ominus A \text{ iff } \models_{t-1}^h A.$$

If points in time are days, for example, $\ominus$ corresponds to an adverbial use of "yesterday", and "John sees to it that yesterday Joan took out the trash" has the form $STIT\alpha \ominus A$.

Provided that A satisfies the *extensionality* condition, one can prove :

$$STIT\alpha \ominus A \text{ is equivalent to } \ominus STIT\alpha A.$$

In other words, for every history h and time t in any model :

$$\models_t^h STIT\alpha \ominus A \text{ iff } \models_t^h \ominus STIT\alpha A.$$

[36] "$STIT\alpha A$ is grammatical and meaningful (though perhaps silly) for an *arbitrary* sentence A" and "$STIT\alpha A$ is always agentive in α, regardless of its complement" (see section 2, above).

[37] *LFI*, pp. 90 – 93.

Stated in terms of the truth conditions of $STIT\alpha \ominus A$ and $\ominus STIT\alpha A$, this is just to say that the conditions

$$\models^{h'}_{t-1} A \text{ for every history } h' \text{ such that } E^{\alpha}_{t'}(h,h')$$
and
$$\not\models^{h'}_{t-1} A \text{ for some history } h' \text{ such that } h \equiv_{t'} h'$$

hold *for some time* $t' < t$ if and only if they hold *for some time* $t' < t-1$. The proof of this in one direction is immediate, since $t' < t-1 < t$. For the reverse, assume that t' is a time earlier than t satisfying the stated conditions. Then $t' \leq t-1$. If $t' < t-1$ we are done. But if we suppose that $t' = t-1$ we shall reach a contradiction. For then our assumption means that

$$\models^{h'}_{t-1} A \text{ for every history } h' \text{ such that } E^{\alpha}_{t-1}(h,h')$$
and
$$\not\models^{h'}_{t-1} A \text{ for some history } h' \text{ such that } h \equiv_{t-1} h'.$$

Since $E^{\alpha}_{t'}(h,h)$, it follows that $h \equiv_{t-1} h'$, $\models^{h}_{t-1} A$, and $\not\models^{h'}_{t-1} A$. This contradicts the *weak* and hence the stronger *extensionality* condition.

The virtue of this equivalence is that one may argue that some sentences of the form $STIT\alpha \ominus A$ are not peculiar after all, in so far as their meanings are explained in terms of the meanings of sentences that are not *STIT–past*. "John sees to it that yesterday Joan took out the trash" proves to have the same meaning as "John saw to it yesterday that Joan took out the trash".

But there exist *STIT–past* sentences that are not equivalent to the corresponding *past–STITs*. For example, let A be an atomic sentence, i.e. one containing no operators. Then :

$$STIT\alpha\langle \mathsf{P}\rangle A \text{ is not equivalent to } \langle \mathsf{P}\rangle STIT\alpha A.$$

In particular, $STIT\alpha\langle \mathsf{P}\rangle A$ does not imply $\langle \mathsf{P}\rangle STIT\alpha A$. The diagram in figure 4 illustrates a model substantiating this contention. We stipulate that histories h and h_1 in the future cone emanating from (h, t_0) are choice equivalent, whereas h_2 is choice equivalent only to itself. So with A true at (h, t_1) and (h_1, t_2), and false everywhere else, $\langle \mathsf{P}\rangle A$ holds at (h, t) and (h_1, t) but not (h_2, t) — and hence, with a witness at (h, t_0), $STIT\alpha\langle \mathsf{P}\rangle A$ holds at (h, t). The failure of A at (h, t_2), (h_1, t_1), and (h, t_0) means that $STIT\alpha A$ is false at (h, t_0), (h, t_1), and (h, t_2) — and therefore $\langle \mathsf{P}\rangle STIT\alpha A$ is false at (h, t).[38]

[38]Notice that the *extensionality* condition is not involved in the example.

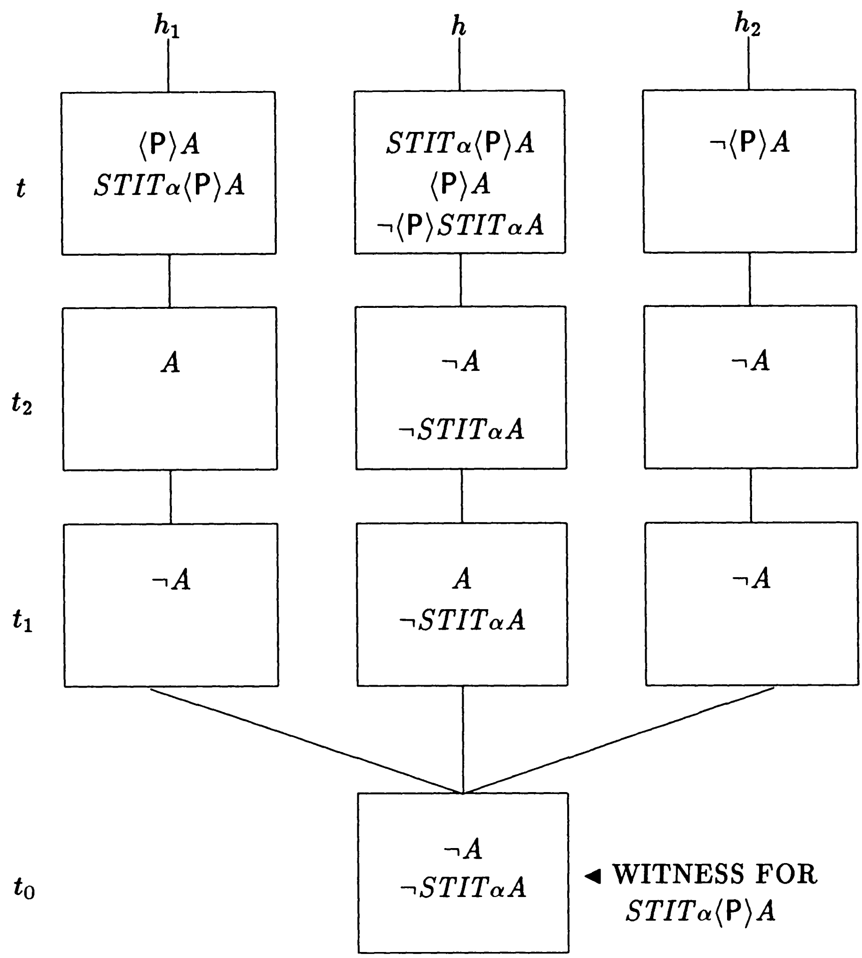

Fig. 4

Thus in *STIT* theory it is possible for an agent to see to something's being so in the past without ever in the past seeing to it that it is so. In terms of responsibility, choice, or however the agency of *sees to it that* is construed, this result blemishes this analysis of *STIT*.

14. Chain witnesses

The problem about *STIT–past* sentences stems from the temporal gap between an agent's choice point and the time of agency itself. If there were no gap, if witnesses were always as close as possible to the moment of agency, examples like the one given in the preceding section could not exist.[39]

Belnap notices the awkwardness of witness–agency temporal gaps in connection with other problematic examples. His solution is to generalize the notion of a witness so that chains of moments can function as witnesses.[40]

The idea is simple enough: a *chain* is a sequence of moments in a single history, i.e. a set of moments all of whose members are comparable under the relation of being earlier than. Within the modeling of *LFI*, a chain can be identified with a set of history–time pairs (h, t) for a fixed history h, *the history of* the chain. For brevity we speak of a time t being in a chain c, meaning that t is part of a history–time pair in c, and we say that chain c is earlier than a time t just when all its times are earlier than t. Let us say also that chains are *comparable* just in case their histories are the same; that one of two comparable chains is *earlier than* the other just in case the first is earlier than some time in the second; and that comparable chains are *co-terminal* if and only if neither is earlier than the other.[41]

When we move from point witnesses to chain witnesses, chains replace times in the indices for choice equivalence relations. Belnap uses the old, point-wise choice equivalence relations to define the new ones. Here is his definition of histories h and h' choice equivalent for agent α with respect to a chain c, fashioned in terms of the modeling of *LFI*:

$$E^{\alpha}_{c}(h,h') \text{ iff } \begin{array}{l} (1) \text{ there is a time } t \text{ in } c \text{ such that } h \equiv_t h' \\ \text{and} \\ (2) \text{ for every time } t \text{ in } c,\ E^{\alpha}_{t}(h,h'). \end{array}$$

Clause (1) ensures that the histories share some portion of the chain, though they may branch off from it at different time; (2) means that the histories are choice equivalent in the old sense with respect to every time in the chain.[42]

[39]To see this, look again at figure 4 and note that if the witness were as close as possible to (h, t) then it would be impossible both for $\langle \mathsf{P} \rangle A$ to be true there and for $\neg\langle \mathsf{P} \rangle A$ to hold at the counter (h_2, t).

[40]One problem has to do with continuing to (choose to) run a mile in ten minutes. Another involves the idea of *refraining* and implication from *refrains from refraining to* to *sees to*. See [1], "Stit pictures", and [2], "Semantics for *stit*". Belnap and Perloff anticipate the development of chain witnesses in [3] and [4] ("Semantics"), writing that "there is room to enlarge our account to include cases in which $[STIT\alpha A]$ is true on the basis of an interval of choices rather than on the basis of a single choice moment".

[41]Note that chains may lack first or last members, and they may be "gappy". The fullest account is in [2], where Belnap details various important properties of histories, chains, and chain witnesses.

[42]The definition is in [2], "Witness of *stit* by chains". By clause (1) if the histories share a point they are identical from then onward into the past. Belnap remarks the awkwardness

In terms of histories, times, agents, and chains, revised truth conditions for *STIT* sentences can be phrased as follows :

$\models_t^h STIT\alpha A$ iff there is a (nonempty) chain c earlier than t such that
(1) $\models_t^{h'} A$ for every history h' such that $E_c^\alpha(h,h')$
and
(2) for every time t' in c there is a history h' such that
$h \equiv_{t'} h'$ and $\not\models_t^{h'} A$.

Thus chain witnesses are earlier than the time of evaluation, and the *positive condition* is unchanged apart from the introduction of a chain in place of a moment or time. The *negative condition* now requires that counters exist with respect to every moment in the chain.[43]

As in the case of point witnesses, Belnap uses the stronger notion of settled truth in stating the *positive* and *negative conditions*. But, again, mere truth will suffice. Let us substantiate this briefly by proving that where c is earlier than t the new *positive condition* above implies its settled counterpart :

$$\text{(S1)} \qquad ||\!\!\models_t^{h'} A \text{ for every history } h' \text{ such that } E_c^\alpha(h,h').$$

Given c earlier than t and *positive condition* (1), assume also that $E_c^\alpha(h,h')$. To show that A is settled true at (h',t), we posit a history h'' such that $h' \equiv_t h''$ and argue that A is true at (h'',t). By definition of E_c^α, there is a time t' in c such that $h \equiv_{t'} h'$, and for every time t' in c, $E_{t'}^\alpha(h,h')$. Because $h' \equiv_t h''$, we may use *no choice between undivided histories* to infer that $E_{t'}^\alpha(h',h'')$ for every time $t' < t$, and hence (since c is earlier than t) for every time t' in c. So by transitivity of the time-indexed choice equivalence relations, $E_{t'}^\alpha(h,h'')$ for every time t' in c. But again, $h' \equiv_t h''$, which means that c contains (h'',t') at every time t' that it contains (h',t'), and so $h \equiv_{t'} h''$. Therefore $E_c^\alpha(h,h'')$ — and hence by (1), A is true at (h'',t).

Unlike their counterparts, chain witnesses are not unique, although results in the spirit of the *witness identity lemma* and its corollaries can be obtained. The basic result is :

If A implies B and $\models_t^h STIT\alpha A$ and $\models_t^h STIT\beta B$, then no chain witness for $STIT\alpha A$ is earlier than any chain witness for $STIT\beta B$.

---

of maintaining terminology here inasmuch as chain-indexed choice equivalence relations are no longer transitive (he gives a counterexample), although they are still reflexive and symmetric.

[43] Compare Belnap's definition e.g. in [2], "Semantics for *stit*".

As a corollary, witness chains for *STIT* sentences true at a history–time pair are all co–terminal when the sentences' complements are equivalent. Therefore, as Belnap points out, for a *STIT* sentence true at a history–time pair all witnesses are co–terminal.[44]

The logic of *STIT* is somewhat altered under the revised truth conditions. Although it is clear that the rule RE persists, as do T, $\overline{\mathsf{N}}$, and the disavowal of M and the rule RM, the status of other principles is not so obvious.[45] But what concerns us is the fact that the *STIT–past* problem remains. For chains may be singletons, and this is of course enough to permit replication of the *STIT–past* example in figure 4 above.

The revised theory provides a framework, however, for avoidance of problematic instances of $STIT\alpha\langle \mathsf{P}\rangle A$. It would be enough, for example, to require that each chain witness approach the point of agency as closely as possible, thereby eliminating the temporal gaps between choice and agency that permit production of troublesome examples. Imposition of such a *nearly now* requirement for chain witnesses would mean that it is always *at least the immediate past* that is relevant to whether or not a *STIT* sentence holds at a time in a history: no matter how early a choice is initiated, it continues up to the very moment of agency.

This *nearly now* requirement is just a suggestion; there may be other ways of explaining away *STIT–pasts.* But although he finds *nearly now* witnesses congenial in the resolution of certain problems, Belnap does not make *nearly now* a universal requirement.[46]

---

[44] Proofs of these results are like those for the corresponding point witness propositions, and rely similarly on the *no choice between undivided histories* condition; we forbear giving them here. In [2], "Witness of *stit* by chains", Belnap canvasses possibilities for the make–up of chain witnesses. He identifies certain maximal chains as candidates for what he calls the "locus of choice", but says that the uniqueness of these witnesses is "of largely technical interest". In the direction of minimality, he notes that deleting vacuous choice points does not disturb witnesshood, and that any "proper initial (i.e., early stretch) of a witness can be subtracted at will", and the residue will remain a witness. In a limiting case of the latter operation, if the chain has a latest history–time pair or moment we arrive at a singleton or its element, a point witness, which Belnap calls a "momentary" witness.

[45] In[2], "Semantics for *stit*", Belnap promises to "elaborate [elsewhere] on the provable properties of *stit*" that result when chain witnesses are utilized. Certain chain witnesses allow for "Busy Choosers", which enable Belnap to avoid his ten–minute mile problem and to defeat the inference from *refrains from refraining to* to *sees to* ([2], "Semantics for *stit*"; see also [1], "*Stit* pictures").

[46] In [2] Belnap refers to his "definition of what it means for a typically nonterminating *chain* of moments jointly to witness the truth" of a *STIT* sentence ("Abstract"), and he mentions "the quest for an understanding of the witness of a *stit* by a possibly unending sequence of choices" ("Witness of *stit* by chains"). It is *nearly now* witnesses that permit the "Busy Choosers" Belnap uses to show that *sees to* does not follow from *refrains from refraining to*, at least absent a restriction that precludes "infinitely many nonvacuous

It is agreeable to observe that a *nearly now* constraint would bring the choice relations into consonance with the instigative relations in *LFI*. For instigative alternatives are required from the outset to have *entirely* the same past history at the time of agency, and it is features of this past that determine which historical alternatives are also instigative. Although recent features of the past are presumably the most significant, it is hard to think why in general the whole past history should not be relevant to the question of what an agent sees to at any given point in time.

15. Concluding remarks

In this paper I have endeavored to set out what I believe are the important aspects of the theories of agency of Belnap and Perloff, and I have criticized the theories at several points. It will be evident, first of all, that I believe that the *negative condition*, and indeed any such negative component of truth conditions for agency sentences, does not form a proper part of the meaning of *sees to it that*. It may be that controversial 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 *STIT* 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 $\Delta\alpha A \wedge \neg \boxdot A$, to see how well its logic replicates that of $STIT\alpha A$.

My objections to the *negative condition* are of course strongly linked to concerns about the principles N and $\overline{\mathsf{N}}$, on the one hand, and closure under consequence (M and the rule RM), on the other. The logic of Δ in *LFI* is defective in so far as it admits N as valid, and the logic of *STIT* is likewise flawed by the presence of $\overline{\mathsf{N}}$. Both are wrong, and I think a satisfactory logic of agency should do without either. It seems to me that the basic logic of agency is *regular*, not normal, which is to say that RE, M, and C all hold and N (likewise $\overline{\mathsf{N}}$) does not. Semantics of a "binary relational" kind are available for such logics; the question is how to interpret these philosophically.[47]

Abandonment of $\overline{\mathsf{N}}$ at least makes M and RM possible, as noted, and I find "Good Samaritan" counterexamples wholly unconvincing. But I have

choices between two given moments" ([2], "Semantics for *stit*"; see also [1], "*Stit* pictures").

[47] Regular logics thus have RM and RR and if non-normal lack RK and RN. See [8], pp. 234 - 245. Essentially their "binary relational" semantics involves a stipulation that points of evaluation be in an appropriate sense "normal". See e.g. [14], pp. 13 - 27, or [8], p.75. It is not difficult to think of philosophical interpretations of this formal constraint, but I would prefer to pursue this on another occasion.

no strong arguments in favor of these principles. I can say only that their counterparts in natural language ring true — and without M and RM there is not much to the logic of agency!

That $STIT\alpha STIT\beta A$ is false whenever α and β are different agents is an astonishing result, and disturbing. Why should one agent not be able to see to it that another sees to something, e.g. in a chain of command? Of course *something happens* is not alone responsible, abetted as it is by the *negative condition*. But even for those who find the other parts of Belnap and Perloff's account acceptable, the correctness of the *something happens* condition must be doubted, and a more searching analysis of the question of *facere per alium* be desired.

Finally, the examples of *STIT–past* sentences constitute in my judgment an extraordinary problem, both for Belnap and Perloff's original, point witness theory and for the revision in terms of chain witnesses. I have to emphasize my agreement with Belnap and Perloff that such sentences must be accepted as syntactically well formed, meaningful, and agentive. Yet these sentences demand an analysis that avoids the conclusion that the agent can sometimes see to something that is already so. A *nearly now* requirement would solve the problem. But absent such a preventive measure, or some otherwise satisfactory interpretation of *STIT–past*, I think this is a defect in the theories.

At the end of a survey of the modal logic of agency, Belnap says of a work of C. L. Hamblin's that "no one ought to try to move deeply into any part of the theory of agency without reading this important book".[48] I would like to echo this sentiment by saying, in closing, that no one *can* move deeply into any part of the theory of agency without reading Nuel Belnap's and Michael Perloff's important papers. Collectively they form a new exciting point of departure for our understanding of this subject.

References

[1] NUEL D. BELNAP, *Backwards and forwards in the modal logic of agency*, University of Pittsburgh, 3 June 1990. **Philosophy and phenomenological research**, forthcoming. Belnap presented an earlier version of this paper under the title "The backwards and forwards of the modal logic of agency" at a Pacific Division meeting of the American Philosophical Association in March 1989.

[2] NUEL D. BELNAP, *Before refraining: concepts for agency.* University of Pittsburgh, 21 September 1989. **Erkenntnis**, forthcoming.

[48][1], "Mini-history". The book is [11].

[3] NUEL D. BELNAP and MICHAEL PERLOFF, *Seeing to it that: a canonical form for agentives*, **Theoria** 54 (1989), pp. 175 - 199. In [1] Belnap notes that "the informal semantic account of *stit* is garbled in this paper; the account is correct in [4]".

[4] NUEL D. BELNAP and MICHAEL PERLOFF, *Seeing to it that: a canonical form for agentives*, in **Knowledge, representation and defeasible reasoning**, H. E. Kyburg, Jr., R. P. Loui, and G. N. Carlson (eds), Kluwer Academic Publishers, Dordrecht 1990, pp. 167 - 190. This is a corrected version of [3], originally read at the 1988 meeting of the Society for Exact Philosophy, in Rochester, New York.

[5] JOHN P. BURGESS, *Basic tense logic*, in **Handbook of philosophical logic. Vol.ii: extensions of classical logic**, Dov M. Gabbay and Franz Guenthner (eds), D. Reidel Publ. Co., Dordrecht 1984, pp. 89 - 133.

[6] BRIAN F. CHELLAS, **The logical form of imperatives**, Perry Lane Press, Stanford 1969. (Ph.D. dissertation Stanford University 1969.)

[7] BRIAN F. CHELLAS, *Imperatives*, **Theoria** 37 (1971), pp. 114 - 129.

[8] BRIAN F. CHELLAS, **Modal logic: an introduction**, Cambridge University Press, Cambridge and New York 1980; reprinted with corrections 1988.

[9] DOV M. GABBAY, **Investigations in modal and tense logics with applications to problems in philosophy and linguistics**, D. Reidel Publ. Co., Dordrecht 1976.

[10] DOV M. GABBAY and FRANZ GUENTHNER, *A note on many-dimensional tense logics*, in **Philosophical essays dedicated to Lennart Åquist on his fiftieth birthday**, T. Pauli (ed.), University of Uppsala, 1982, pp. 63 - 70.

[11] C. L. HAMBLIN, **Imperatives**, Basil Blackwell, Oxford and New York 1987.

[12] J. A. W. KAMP, *Formal properties of 'now'*, **Theoria** 37 (1971), pp.227 - 273.

[13] STORRS MCCALL, *Objective time flow*, **Philosophy of science** 43 (1976), pp. 337 - 362.

[14] KRISTER SEGERBERG, **An essay in classical modal logic**, University of Uppsala: Philosophical Studies, 1971.

[15] RICHMOND H. THOMASON, *Indeterminist time and truth-values gaps*, **Theoria** 36 (1970), pp. 264 - 281.

[16] RICHMOND H. THOMASON, *Combination of tense and modality*, in **Handbook of philosophical logic. Vol.ii: extensions of classical logic**, Dov M. Gabbay and Franz Guenthner (eds), D. Reidel Publ. Co., Dordrecht 1984, pp. 135 - 165.

[17] FRANK VLACH, *Now and then: a formal study in the logic of tense anaphora.* Ph.D dissertation, University of California at Los Angeles 1973.

DEPARTMENT OF PHILOSOPHY
THE UNIVERSITY OF CALGARY
CALGARY, ALBERTA
CANADA T2N 1N4

Received June 6, 1991