Action Types in Stit Semantics

John Horty

Eric Pacuit

University of Maryland

Version of: May 31, 2017

Abstract

Stit semantics grows out of a modal tradition in the logic of action that concentrates on an operator representing the agency of an individual in seeing to it that some state of affairs holds, rather than on the actions the individual performs in doing so. The purpose of this paper is to enrich stit semantics, and especially epistemic stit semantics, by supplementing the overall framework with an explicit treatment of action types. We show how the introduction of these new action types allows us to define a modal operator capturing an epistemic sense of agency, and how this operator can be used to express an epistemic sense of ability.

Contents

1 Introduction

Stit semantics—originating with a series of papers by Belnap, Perloff, and Xu, culminating in their [4]—grows out of a modal tradition in the logic of action going back to St. Anselm, but with more recent contributions by, among others, Anderson, Chellas, Fitch, Kanger, Lindahl, Pörn, and von Kutschera.[1] It is characteristic of this tradition to focus on a modal operator representing the agency of an individual in bringing it about that—or *seeing to it that*, hence *stit*—some state of affairs holds, rather than on the actions that the individual carries out in doing so. In a recent survey, Lindström and Segerberg describe the work in this tradition as a as a "logic of action without actions," writing that:

> No author in the Anselm-Kanger-Chellas line up through Belnap—Davidson belongs to a different tradition—has countenanced the existence of actions in logic: action talk, yes; ontology of action, no.[2]

In fact, and as Lindström and Segerberg go on to note, one of the present authors, working somewhat later in the tradition of stit semantics, does speak explicitly of actions in [12], where a deontic operator is defined in terms of a preference ordering on the actions available to an individual. The actions discussed in that book, however, were action tokens—particular, concrete actions, each occurring at a single point in space and time. There were no general, repeatable kinds of actions, or action types, such as the action type of "opening a window," for example; there were only particular openings of particular windows, with nothing to group them together as actions of the same kind.

[1] A brief history of the subject, with references to the works of these writers and others, can be found in Segerberg [23].

[2] Lindström and Segerberg [18, p. 1199].

The purpose of this paper is to enrich stit semantics, and especially epistemic stit semantics, by supplementing the overall framework with an explicit treatment of action types, in addition to the action tokens that were already present. But we do not do this simply for the sake of doing it, or because the neglect of action types seems like a defect from an external perspective. Instead, working from a perspective internal to stit semantics, we motivate the introduction of action types by showing how they help us to represent an important concept: the epistemic sense of ability.

The paper is organized as follows: The next section provides a summary of basic stit semantics, leading up to the definition of a standard stit operator. Section 3 considers the standard treatment of ability within the theory, shows that this treatment fails to capture the epistemic sense of this concept, and argues that simply supplementing the basic theory with an epistemic operator will not work either. Section 4 introduces the new framework of *labeled stit semantics*, where the label assigned to an action token represents the action type of which that token is an instance; within this framework, a new epistemic stit operator is defined, which then allows for an analysis of the epistemic sense of ability. Section 5 discusses some logical points concerning the epistemic stit operator introduced here, as well as connections with the existing literature.

2 A review of stit semantics

2.1 Branching time

Stit semantics is cast against the background of a theory of indeterministic time, first set out by Prior [21] and developed in more detail by Thomason [25], according to which moments

are ordered into a treelike structure, with forward branching representing the indeterminacy of the future and the absence of backward branching representing the determinacy of the past.

This picture leads to a notion of *branching time frames* as structures of the form $\langle \mathit{Tree}, < \rangle$, in which *Tree* is a nonempty set of moments and $<$ is a strict partial ordering of these moments without backward branching: for any m, m', and m'' from *Tree*, if $m' < m$ and $m'' < m$, then either $m' = m''$ or $m'' < m'$ or $m' < m''$. A maximal set of linearly ordered moments from *Tree* is a *history*, representing some complete temporal evolution of the world. If m is a moment and h is a history, then the statement that $m \in h$ can be taken to mean that m occurs at some point in the course of the history h, or that h passes through m. Because of indeterminism, a single moment might be contained in several distinct histories. We let $H^m = \{h : m \in h\}$ represent the set of histories passing through m; and when h belongs to H^m, we speak of a moment/history pair of the form m/h as an *index*.

A *branching time model* is a structure that supplements a branching time frame with a valuation function v mapping each propositional constant from some background language into the set of m/h indices at which, intuitively, it is thought of as true. If we suppose that formulas are formed from truth functional connectives as well as the usual temporal operators P and F, representing past and future, the satisfaction relation $\models$ between indices and formulas true at those indices is defined as follows.

Definition 1 (Evaluation rules: basic operators) Where m/h is an index and v the evaluation function from a branching time model $\mathcal{M}$,

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

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

In addition to the usual temporal operators, the framework of branching time allows us to define the concept of historical necessity, along with its dual concept of historical possibility: the formula $\Box A$ is taken to mean that A is historically necessary, while $\Diamond A$ means that A is still open as a possibility. The intuitive idea is that $\Box A$ is true at some moment if A is true at that moment no matter how the future turns out, and that $\Diamond A$ is true if there is still some way in which the future might evolve that would lead to the truth of A. The evaluation rule for historical necessity is straightforward.

Definition 2 (Evaluation rule: $\Box A$) Where m/h is an index from a branching time model $\mathcal{M}$,

- $\mathcal{M}, m/h \models \Box A$ if and only if $\mathcal{M}, m/h' \models A$ for each history $h' \in H^m$.

And historical possibility can then be characterized in the usual way, with $\Diamond A$ defined as $\neg\Box\neg A$.

The notion of historical necessity can be registered in the metalanguage by defining a formula A as *settled true* at a moment m from a model $\mathcal{M}$ just in case $\mathcal{M}, m/h \models \Box A$; likewise A can be defined as *settled false* just in case $\mathcal{M}, m/h \models \Box\neg A$.

In branching time, the set of possible worlds accessible at a moment m can be identified with the set H^m of histories passing through that moment; those histories lying outside

of H^m are taken to represent worlds that are no longer accessible. The propositions at m can thus be identified with sets of accessible histories, subsets of H^m. And the particular proposition expressed by a sentence A at a moment m in a model $\mathcal{M}$ can be identified with the set $|A|^m_{\mathcal{M}} = \{h \in H^m : \mathcal{M}, m/h \models A\}$ of histories from H^m in which that sentence is true. Here and elsewhere, we will omit reference to the background model when context allows, writing $|A|^m$, for example, to refer to the proposition expressed by A in some model that can be identified by the context, or in an arbitrary model.

2.2 The *stit* operator

Within stit semantics, the idea that an agent α sees to it that A is taken to mean that the truth of A is guaranteed by an action performed by α. In order to capture this idea, we must be able to speak of individual agents, and also of their actions; and so the basic framework of branching time is supplemented with two additional primitives.

The first is simply a set *Agent* of agents, individuals thought of as acting in time. The second is a device for representing the possible constraints that a particular agent is able to exercise upon the course of events at a given moment, the actions or choices open to the agent at that moment. These constraints are encoded through a function *Choice*, mapping each agent α and moment m into a partition $Choice^m_\alpha$ of the set H^m of histories through m. The idea is that, by acting at m, the agent α is able to determine a particular one of the equivalence classes from $Choice^m_\alpha$ within which the history to be realized must then lie, but that this is the extent of the agent's influence.

If K is a choice cell from $Choice^m_\alpha$, one of the equivalence classes specified by the partition, we speak of K as an action—or more precisely, an *action token*—available to the agent α

at the moment m, and we say that α *performs* the action token K at the index m/h just in case h is a history belonging to K. We let $Choice^m_\alpha(h)$ (defined only when $h \in H^m$) stand for the particular equivalence class from $Choice^m_\alpha$ that contains the history h; $Choice^m_\alpha(h)$ thus represents the particular action token performed by the agent α at the index m/h.

Apart from specifying, for each agent, a partition of the histories through each moment, the *Choice* function is subject to two further requirements. The first is a condition of *independence of agents*, which says, roughly, that, at any given moment, any selection of actions tokens by different agents must be consistent, or nonempty.[3] The second requirement stipulates that the choices available to an agent at a moment should not allow a distinction between histories that do not divide until some later moment. Let us say that two histories are undivided at m whenever they share a moment that is properly later than m. The requirement of *no choice between undivided histories* can then be expressed as the condition that, for each agent α, any two histories that are undivided at m must belong to the same choice cell from $Choice^m_\alpha$.

With these new primitives, a *stit frame* can be defined as a structure of the form

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

supplementing a branching time frame with the additional components *Agent* and *Choice*, as specified above, and a *stit model* as a model based on a stit frame.

We can now define a standard stit operator, written [... *stit*: ...] and allowing for statements of the form [α *stit*: A], with the intuitive meaning that α sees to it that A. Such a statement is defined as true at an index m/h just in case the action token performed by

[3] A more precise definition of this independence requirement can be found in Section 2.4 of [12].

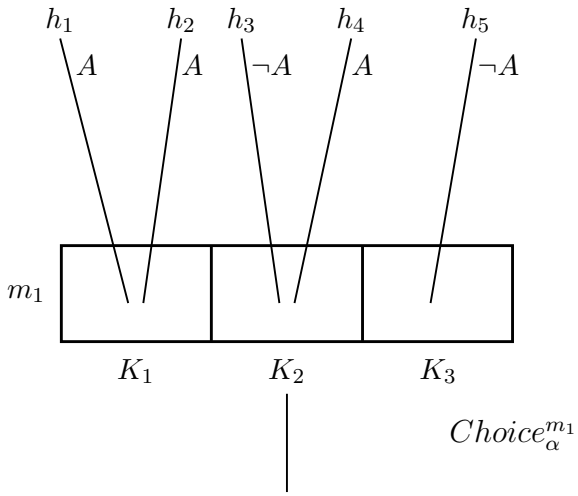

Figure 1: [α *stit*: A] true at m_1/h_1

α at that index guarantees the truth of A. Formally, we can say that some action token K available to an agent at the moment m guarantees the truth of A just in case A holds at m/h for each history h from K—just in case, that is, $K \subseteq |A|^m$. Since the action token performed by α at the index m/h is $Choice^m_\alpha(h)$, our semantic analysis can be captured through the following evaluation rule.[4]

Definition 3 (Evaluation rule: [α *stit*: A]**)** Where α is an agent and m/h an index from a stit model $\mathcal{M}$,

- $\mathcal{M}, m/h \models$ [α *stit*: A] if and only if $Choice^m_\alpha(h) \subseteq |A|^m_{\mathcal{M}}$.

These various definitions are illustrated in Figure 1, where $Choice^{m_1}_\alpha = \{K_1, K_2, K_3\}$, with $K_1 = \{h_1, h_2\}$, $K_2 = \{h_3, h_4\}$, and $K_3 = \{h_5\}$.[5] Here, the statement [α *stit*: A] is true at the index m_1/h_1, for example, since $Choice^{m_1}_\alpha(h_1) = K_1$ and $|A|^{m_1} = \{h_1, h_2, h_4\}$, so that

[4]Those familiar with stit logics will recognize this particular operator as the "Chellas stit," first introduced into stit logics by Horty and Belnap [13], but drawing on ideas from Chellas [10].

[5]A convention for interpreting figures: when a formula is written next to some history emanating from

$Choice_\alpha^{m_1}(h_1) \subseteq |A|^{m_1}$. But $[\alpha\ \mathit{stit}\colon A]$ is not true at m_1/h_4, since $Choice_\alpha^{m_1}(h_4) = K_2$, so that we do not have $Choice_\alpha^{m_1}(h_4) \subseteq |A|^{m_1}$. Even though the statement A itself happens to hold at m_1/h_4, the action token K_2 that is performed by α at this index does not guarantee the truth of A.

3 Ability and knowledge

3.1 Ability

There are several ways of motivating the introduction of action types, in addition to tokens, into stit semantics, but we concentrate here on a motivation deriving from problems with the stit characterization of personal ability.

This notion of ability must be distinguished, of course, from that of mere possibility: even though it is possible for it to rain tomorrow, no agent has the ability to see to it that it will rain tomorrow. Nevertheless, it has been suggested in the stit literature that what an agent is able to do can reasonably be identified with what it is possible that the agent does. In particular, Horty and Belnap [13] proposed that a formula of the form

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

carrying the literal meaning that it is possible for the agent α to see to it that A, can usefully be taken to expresses the claim that α has the ability to see to it that A. The general idea behind this proposal is that α has the ability to see to it that A just in case there is some action—in this case, an action token—available to α that guarantees the truth of A.

a moment, the formula should be taken as true at that moment/history pair. Thus, A should be taken as true at m_1/h_1 in Figure 1, for example, and $\neg A$ as true at m_1/h_3.

It may appear that this suggestion is vulnerable to the well-known argument advanced by Kenny, in [16] and [17], that ability cannot be analyzed using the techniques of modal logic—or as he puts it, referring to ability as a "dynamic modality," that "ability is not any kind of possibility ... dynamic modality is not a modality."[6] Kenny's argument centers around statements of the form

$$A \supset \mathit{Can}(A),$$

$$\mathit{Can}(A \vee B) \supset (\mathit{Can}(A) \vee \mathit{Can}(B)),$$

where *Can* represents a possibility operator developed within the usual framework of modal logic. The first of these statements is valid in any reflexive modal logic, and the second is valid in any normal modal logic, but Kenny argues that neither should be taken to characterize the logic of ability. As a counterexample to the first, he considers the case in which a poor darts player throws a dart and happens to hit the bull's eye. Although this shows that it is possible for the darts player to hit the bull's eye, it does not seem to establish that she has the ability to do so. As a counterexample to the second, Kenny imagines another darts player whose skill is sufficient to guarantee only that the dart hits the dart board, but who has no further control of the dart beyond that. Since any dart that hits the dart board must land either in the top half or in the bottom half, this player has the ability to hit either the top half or to hit the bottom half of the dart board; but since the player has no further control, she does not have the ability to hit the top half, or the ability to hit the bottom half.

Even though the analysis of ability proposed by Horty and Belnap is developed within a modal framework, of the sort that Kenny objects to, the analysis escapes his objections, since

[6]See Kenny [17, p. 226].

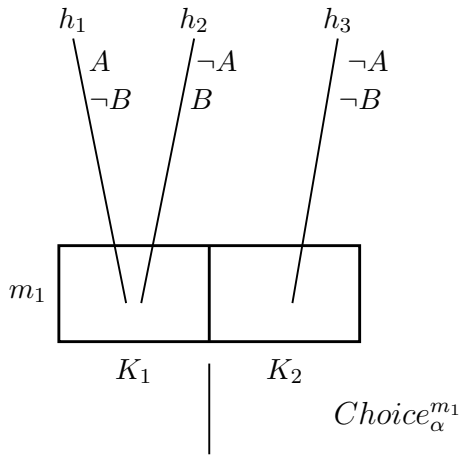

Figure 2: The darts examples

it is not historical possibility alone that is taken to represent ability, but rather a combination of historical possibility together with a stit operator. This combination, it turns out, fails to validate the formulas of the form that Kenny was concerned with: both

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

$$\Diamond[\alpha \; \textit{stit}\colon A \vee B] \supset (\Diamond[\alpha \; \textit{stit}\colon A] \vee \Diamond[\alpha \; \textit{stit}\colon B])$$

can be falsified. A joint countermodel, based on Kenny's darts stories, is provided in Figure 2, where we can imagine that the agent α must choose at the moment m_1 between the action token K_1 of throwing the dart and the action token K_2 of refraining. To interpret Kenny's first story, we take A to mean that the dart will hit the bull's eye (and we ignore the sentence B). Here, if α chooses to throw the dart and things evolve along the history h_1, then the dart will hit the bull's eye. But this is not a proposition whose truth the agent has the ability to guarantee: although A is true at m_1/h_1, the formula $\Diamond[\alpha \; \textit{stit}\colon A]$ is not. For the second story, we take A to mean that the dart will hit the top half of the dart board, and B to mean that the dart will hit the bottom half. Since, by performing the action token K_1,

the agent is able to guarantee that the dart will hit either the top half or the bottom half of the dart board, the formula $\Diamond[\alpha \; \mathit{stit}\colon A \vee B]$ is settled true at m_1. But both $\Diamond[\alpha \; \mathit{stit}\colon A]$ and $\Diamond[\alpha \; \mathit{stit}\colon B]$ are settled false, since no action token available to α guarantees that the dart will hit the top half of the dart board, and no action token available to α guarantees that the dart will hit the bottom half.

The proposed analysis of ability, then, allows for a sensible response to Kenny's arguments. It bears, in addition, clear relations, explored in [13], to the logic of ability developed by Brown [9], and also to treatments of ability developed in the coalition logic introduced by Pauly [20] as well in the alternating-time temporal logic due to Alur, Henzinger, and Kupferman [2]; these later relations were first explored by Broersen, Herzig, and Troquard in a series of papers beginning with [7] and [8].

We still feel that this proposal captures one important sense of ability, which we refer to here as the *causal* sense. There is also, however, another important sense of ability that the proposal does not capture, and which we can describe as the *epistemic* sense. To illustrate: suppose a friend places all the cards from a deck face down on a table, and asks you to turn over the Jack of Hearts. Is that something you are able to do? The answer is Yes, in the causal sense of ability. There are, we can suppose, 52 actions available to you—turning over any of the cards. Each of these actions guarantees that some particular card is turned over, and the Jack of Hearts is among them; so one of the actions available to you guarantees that the Jack of Hearts is turned over. But in the epistemic sense, the answer is No. Even if one of the actions available to you happens to guarantee that the Jack of Hearts is turned over, you do not have, in the epistemic sense, an ability to turn that card over unless you also know which of your available actions guarantees the result.

3.2 Knowledge

In order to understand the epistemic sense of ability, it is natural, as a first step, to introduce a knowledge operator into stit logic. This can be done in the standard way.[7] Begin by positing, for each agent α, an equivalence relation $\sim_\alpha$ among the indices from a stit frame, where $m/h \sim_\alpha m'/h'$ is taken to mean that nothing α knows distinguishes m/h from m'/h', or that m/h and m'/h' are epistemically indistinguishable by α. An *epistemic stit frame* can be defined as a structure of the form

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}\rangle,$$

like a stit frame but with the additional component $\{\sim_\alpha\}_{\alpha \in Agent}$, a set containing indistinguishability relations for the various agents from $Agent$; and an *epistemic stit model* can be defined by supplementing an epistemic stit frame with a valuation function. The evaluation rule for an operator of the form K_α, representing knowledge for the agent α, can then be introduced, as usual, through the stipulation that an agent at an index knows whatever holds at all indices indistinguishable from that one.

Definition 4 (Evaluation rule: $\mathsf{K}_\alpha A$) Where α is an agent and m/h an index from an epistemic stit model $\mathcal{M}$,

- $\mathcal{M}, m/h \models \mathsf{K}_\alpha A$ if and only if $\mathcal{M}, m'/h' \models A$ for all m'/h' such that $m'/h' \sim_\alpha m/h$.

Once stit semantics has been supplemented by knowledge operators like these, it is tempting to suppose that the epistemic sense of ability can be represented through some combi-

[7] We say that this treatment of knowledge is standard, and it does follow the usual pattern of epistemic logic; but it was not until Herzig and Troquard [11] that anyone even thought to explore epistemic ideas in the context of stit semantics.

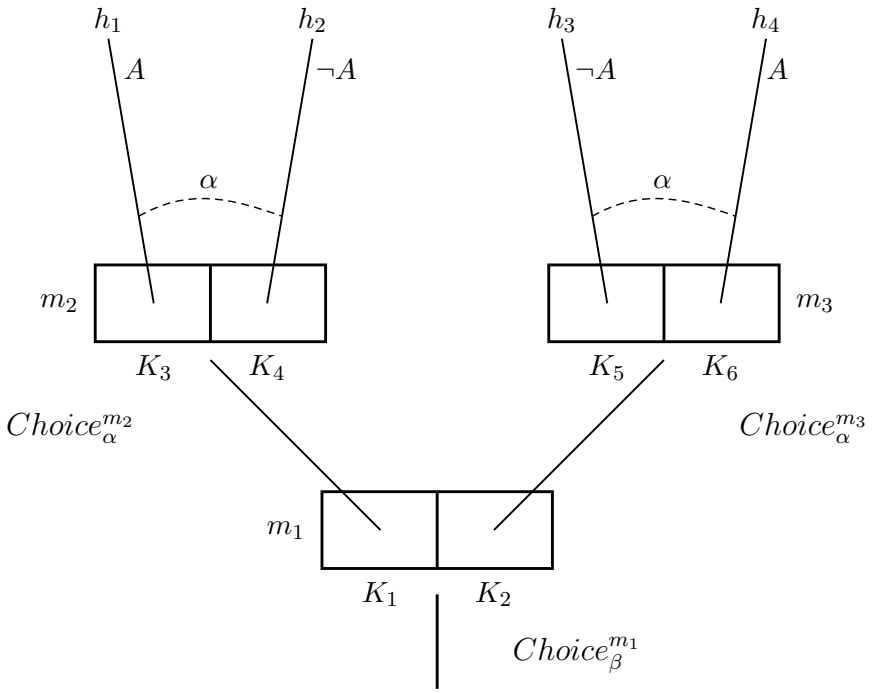

Figure 3: The first coin game

nation of knowledge, impersonal possibility, and agency—perhaps the idea that α has the ability to see to it that A in this epistemic sense should be represented as $\mathsf{K}_\alpha\Diamond[\alpha\ \mathit{stit}\text{: } A]$, or perhaps as $\Diamond\mathsf{K}_\alpha[\alpha\ \mathit{stit}\text{: } A]$. The first of these formulas, taken literally, means that the agent α knows that it is possible that she sees to it that A; the second means that it is possible that α knows that she sees to it that A. We do not feel that either of these analyses is exactly right, for reasons that are best explained with an example.

Consider two simple games, depicted in Figures 3 and 4, in each of which, at an initial moment m_1, the agent β places a coin on the table either heads up, by selecting the action token K_1, or tails up, by selecting K_2. Next, the agent α bets whether the coin was placed heads up or tails up. This action occurs at one of the later moments m_2 or at m_3, depending on the initial choice by β. If β has selected K_1, then α bets heads by selecting K_3 and tails

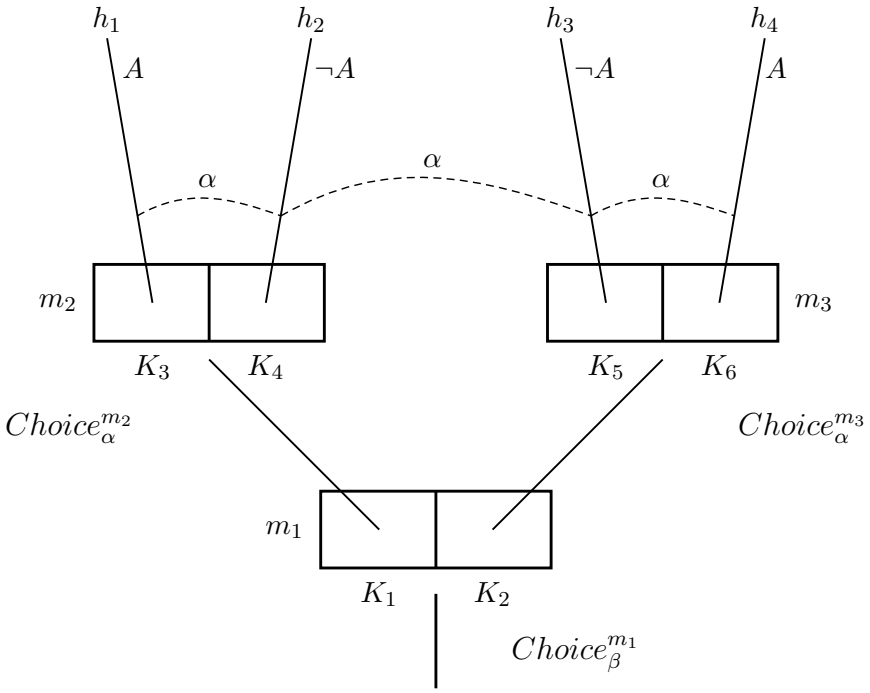

Figure 4: The second coin game

by selecting K_4; if β has selected K_2, then α bets heads by selecting K_5 and tails by selecting K_6. If α bets correctly, she wins. There are four histories: h_1 and h_4 are the histories in which α wins, betting that the coin is heads up when it is, and tails up when it is; h_2 and h_3 are the histories in which α loses. If we let A represent the statement that α wins, then we have A true at m_2/h_1 and m_3/h_4, and false at m_2/h_2 and m_3/h_3.

The two games we consider share this much structure—what we might refer to as their causal structure—but differ in whether or not the agent α knows whether the coin has been placed heads up or tails up, and so must be represented by different epistemic stit models. In the first game, α knows whether the coin has been placed heads up or tails up, and so knows, in effect, whether she is at m_2 or m_3; this game is represented by supposing that the indistinguishability relation $\sim_\alpha$ partitions the indices from m_2 and the indices

from m_3 into the separate, and so distinguishable, equivalence classes $\{m_2/h_1, m_2/h_2\}$ and $\{m_3/h_3, m_3/h_4\}$.[8] In the second game, α does not know whether the coin has been placed heads up or tails up, and so, in effect, does not know whether she is at m_2 or m_3; this game is represented by supposing that the indistinguishability relation $\sim_\alpha$ groups all the indices from m_2 and m_3 together into the single equivalence class $\{m_2/h_1, m_2/h_2, m_3/h_3, m_3/h_4\}$.

Now of course, in both of these games, the agent α has the ability to win in the purely causal sense: the formula $\Diamond[\alpha\ \mathit{stit}\colon A]$ is settled true at both m_2 and m_3, regardless of the agent's knowledge. But when we turn to the epistemic sense of ability, it seems that this knowledge should make a difference. In the first game, where the agent knows whether the coin has been placed heads up or tails up, we would like to say that she does have the ability to win, even in the epistemic sense. But in the second game, where the agent does not know whether the coin has been placed heads up or tails up, we would now like to say that, although the causal ability is still there, she no longer has the ability to win in the epistemic sense. Unfortunately, neither of the two formulas under consideration—that is,

[8] A further convention for interpreting figures: when a history h emanating from a moment m is connected by an α-arc to a history h' emanating from a moment m', it should be understood that $m/h \sim_\alpha m'/h'$, with the $\sim_\alpha$ relation then closed under reflexivity, transitivity, and symmetry. Note also that, in grouping, for example, the indices from K_3 and K_4 together, the indistinguishability relation from this figure suggests that, while the agent knows whether the coin has been placed heads or tails, she does not know whether she is currently betting heads or betting tails. This gap between action and knowledge—exhibited here and in some of the following examples—is a matter that some readers of earlier drafts have objected to, and which we now address directly in Section 5, in terms of the distinction between *ex ante* and *ex interim* knowledge. We hope that readers who are troubled by our treatment of the relation of action to knowledge will bear with us until this later discussion.

neither $\mathsf{K}_\alpha\Diamond[\alpha\ \textit{stit}\text{: } A]$ nor $\Diamond\mathsf{K}_\alpha[\alpha\ \textit{stit}\text{: } A]$—allows us to say both of these things, since, as the reader can verify, the formula $\mathsf{K}_\alpha\Diamond[\alpha\ \textit{stit}\text{: } A]$ is settled true, and $\Diamond\mathsf{K}_\alpha[\alpha\ \textit{stit}\text{: } A]$ is settled false, at both of the moments m_2 and m_3 in both games. What we need, in order to isolate the epistemic sense of ability, is a formula that holds in the first game but fails in the second; but of the two formulas under consideration, one holds, and the other fails, in both.

4 Labeled stit semantics

4.1 Action types

How, then, can we represent the epistemic sense of ability? The general idea underlying the earlier proposal, due to Horty and Belnap, was that an agent α has the ability to see to it that A just in case there is some action—in that case, an action token—available to α that guarantees the truth of A. So let us start with a straightforward epistemic modification of this basic idea. Rather than requiring only that some action token available to α in fact guarantees the truth of A, what we might suppose is that α has the ability to see to it that A in the epistemic sense just in case some available action token is known by α to guarantee the truth of A.

But now, consider a variant of our coin game. Imagine this time that β has two coins, a nickel and a dime, and begins at m_1 by placing either one on the table, either heads up or tails up: we take K_1 as the action token of placing the nickel heads up, K_2 as the action token of placing the dime heads up, K_3 as the action token of placing the nickel tails up, and K_4 as the action token of placing the dime tails up. As before, α then bets whether the coin—whichever coin was placed on the table—was placed heads up or tails up, and wins if

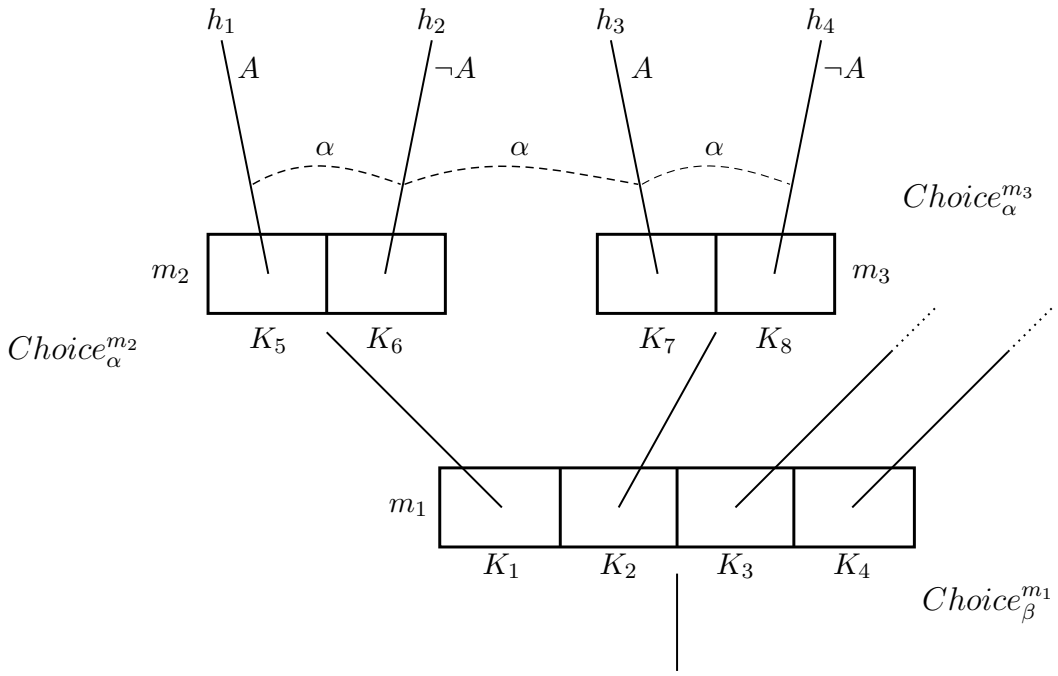

Figure 5: The third coin game

she bets correctly. We consider here only that portion of the game, depicted in Figure 5, in which the coin was in fact placed heads up. In that case, if β has placed the nickel heads up, selecting K_1, then α bets heads by selecting K_5 and tails by selecting K_6; if β has placed the dime heads up on the table, selecting K_2, then α bets heads by selecting K_7 and tails by selecting K_8. Again, we take A as the statement that α wins, so that A is true at m_2/h_1 and m_3/h_3, and false at m_2/h_2 and m_3/h_4. Finally, we assume that, although α does not know which coin was placed on the table, nickel or dime, she does know that whichever coin was placed on the table was placed heads up—that is, we assume that $\sim_\alpha$ groups the indices from m_2 and m_3 together into the single equivalence class $\{m_2/h_1, m_2/h_2, m_3/h_3, m_3/h_4\}$.

In this situation, since α knows that the coin was placed heads up, we would like to say that she has the ability to win even in the epistemic sense: all she needs to do, it seems, is bet heads—that is an action she knows will guarantee a win. But betting heads, in this sense, is not among the action tokens available to α. If α is at m_2, then the action token K_5 is among those available to her, and if she is at m_3, then the action token K_7 is available. These are, however, two different concrete actions, both tokens of what we might call the *action type* of betting heads. It follows that, if the epistemic sense of ability requires that some single action must be known by α to guarantee the truth of A, then this must be the action type of betting heads, not one of its various tokens.

We take this argument to suggest that understanding the epistemic sense of ability involves an appeal to action types, as well as tokens, and turn now to the task of enriching stit semantics with the machinery necessary to treat both tokens and types.

4.2 Basic concepts

We begin by postulating a set $Type = \{\tau_1, \tau_2, \ldots\}$ of action types—general kinds of action, as opposed to the concrete action tokens already present in stit logics. The intuition is that an agent performs a concrete action token at a particular moment by executing one of these action types at that moment. We assume here, for simplicity, that each of the action types belonging to this set is a maximally specific kind of action available to the agent, though it may be interesting to investigate logics in which this assumption is relaxed. We further assume, again for simplicity, that all action types are primitive, though there is no reason to rule out the possibility of complex action types, perhaps specified by a compositional action description language. In contrast to action tokens, action types are repeatable. A robot, for example, might execute the action type of raising its left arm four inches twice during the day, once at the lab in the morning and once at home in the evening, resulting in two concrete action tokens of the same type; a gambler might execute the action type of betting heads in two different games, or at two different points in the same game.

Once action types have been introduced into stit logic, it is most natural to assume that it is the execution of these action types, rather than the performance of concrete action tokens, that falls most directly within the agent's control. This point can be illustrated by returning to our third coin game, from Figure 5, for example. Here, although the agent α knows that some coin has been placed heads up, she does not know whether it was the nickel or the dime—that is, she does not know whether she is at the moment m_2 or the moment m_3. It is hard to see, therefore, how the agent could actually choose to perform either of the concrete action tokens K_5 or K_7, since K_5 is available only at m_2 and K_7 is available only at

m_3. What the agent can do, however, is choose to execute the action type of betting heads, which will then result in the performance of the token K_5 if she is at m_2 and K_7 if she is at m_3.

Formally, the new action types introduced here are related to the action tokens already present in stit semantics through two functions. The first is a partial *execution* function—written, [ ]—mapping each action type τ into the particular action token $[\tau]^m_\alpha$ that results when τ is executed by the agent α at the moment m. Of course, the action token $[\tau]^m_\alpha$ must be one of those available to α at m—that is, we must have $[\tau]^m_\alpha \in Choice^m_\alpha$. The execution function is partial because it seems best, from an intuitive standpoint, to assume that not every action type is available for execution by every agent at every moment.

Just as the execution function maps the action type τ executed by an agent α at a moment m into a particular action token $[\tau]^m_\alpha$ from $Choice^m_\alpha$, we postulate, in addition, a one-one *label* function—written, *Label*—mapping each action token K from $Choice^m_\alpha$ into a particular action type $Label(K)$ from $Type$, where the label assigned to the action token K is, intuitively, the action type that the agent α would have to execute at the moment m in order to perform the action token K. The label function is one-one because it seems best to assume that the execution of different action types leads to the performance of different action tokens.

The interaction between the execution and label functions is governed by two *execution/label constraints*:

If $K \in Choice^m_\alpha$, then $[Label(K)]^m_\alpha = K$,

If $\tau \in Type$ and $[\tau]^m_\alpha$ is defined, then $Label([\tau]^m_\alpha) = \tau$.

The first of these requires that, if K is an action token available to α at m, and K is a token of a particular type $Label(K)$, then the execution of an action of that type by α at m is K itself; the second requires that, if τ is an action type whose execution by α at m is a particular action token $[\tau]^m_\alpha$, then the type of that action token is τ itself.

Our previous definition of the action tokens available to an agent at a moment, as well as our definition of the particular action token performed by an agent at an index, can now be lifted from tokens to types in the natural way. Since $Choice^m_\alpha$ is the set of action tokens available to the agent α at the moment m, we can take

$$Type^m_\alpha = \{Label(K) : K \in Choice^m_\alpha\}$$

as the set of action types available to α at m; and since $Choice^m_\alpha(h)$ is the particular action token performed by α at the index m/h, we can take

$$Type^m_\alpha(h) = Label(Choice^m_\alpha(h))$$

as the action type executed by α at that index.

Putting these various ideas together, we can define a *labeled stit frame* as a structure of the form

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, Type, [\ ], Label\rangle,$$

like an epistemic stit frame, but with the new components $Type$, $[\ ]$, and $Label$ as specified above; a *labeled stit model* results when such a frame is supplemented with a valuation.

4.3 The *kstit* operator

We are now in a position to introduce a new *epistemic* stit operator—written [... *kstit*: ...], and allowing for statements of the form [α *kstit*: A]—distinct from the standard stit operator

presented earlier, which can be described as a *causal* stit. As with our earlier stit statements, a statement of this new form can likewise be interpreted to mean that α sees to it that A, but in a different, epistemic sense. While the earlier statement [α *stit*: A] was taken to mean that α performs an action token guaranteeing the truth of A, what the new statement [α *kstit*: A] means is that α executes an action type that she knows to guarantee the truth of A. More precisely, this statement will be defined as true at an index m/h just in case the action type executed by α at that index guarantees the truth of A at every moment m' participating in an index m'/h' that is indistinguishable from m/h. Since the action type executed by α at the index m/h is $Type^m_\alpha(h)$, the execution of this action type by α at another moment m' is $[Type^m_\alpha(h)]^{m'}_\alpha$. Therefore, the evaluation rule for our new operator is as follows.

Definition 5 (Evaluation rule: [α *kstit*: A]) Where α is an agent and m/h an index from a labeled stit model $\mathcal{M}$,

- $\mathcal{M}, m/h \models [\alpha \ kstit\colon A]$ if and only if $[Type^m_\alpha(h)]^{m'}_\alpha \subseteq |A|^{m'}_{\mathcal{M}}$ for all m'/h' such that $m'/h' \sim_\alpha m/h$.

This definition is beset by an immediate complication, which we can see by noting that the rule begins with an action type $Type^m_\alpha(h)$ executed by the agent α at the index m/h, and then considers the effects arising from an execution of that same action type by the same agent at a different moment m', where m and m' are linked only by participating in indistinguishable indices. In order for this procedure to make sense, and so for the evaluation rule to be well-defined, we need to ensure that the action type executed by the agent at m/h is actually available for execution also at m'. We therefore stipulate that labeled stit frames must satisfy the constraint

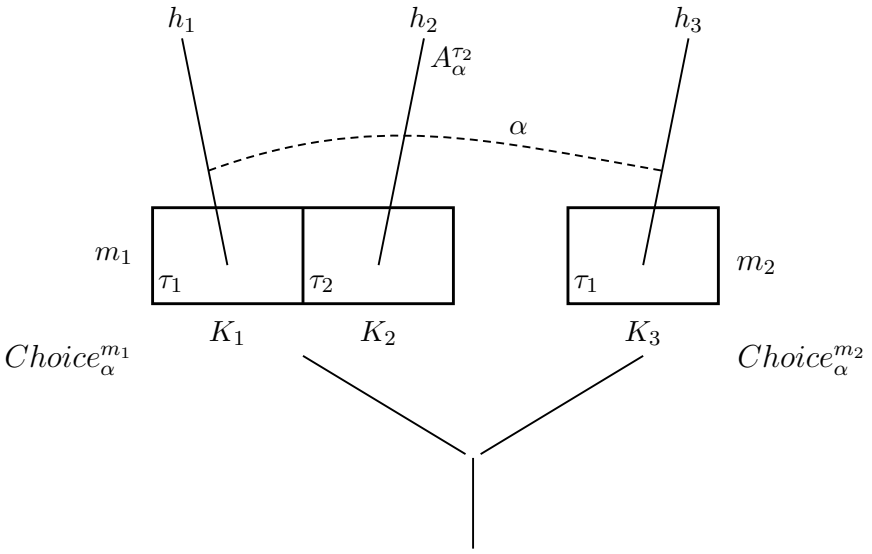

Figure 6: A model satisfying (C2) but not (C1)

(C1) If $m/h \sim_\alpha m'/h'$, then $Type_\alpha^m = Type_\alpha^{m'}$,

which requires that the same action types are available for execution by an agent at any two moments participating in indices indistinguishable for that agent.

Strictly speaking, the constraint (C1) is stronger than necessary. We could have guaranteed that our new operator is well-defined by imposing only the weaker constraint

(C2) If $m/h \sim_\alpha m'/h'$, then $[Type_\alpha^m(h)]_\alpha^{m'}$ is defined,

according to which an action type executed by an agent at some moment participating in an index must be available for execution at all moments participating in indices indistinguishable by that agent from the original. A model illustrating the difference between the constraints (C1) and (C2) is depicted in Figure 6, where $m_1/h_1 \sim_\alpha m_2/h_3$ is the only indistinguishability relation, and where the action tokens K_1 and K_3 belong to the action type τ_1, while the token

K_2 belongs to the type τ_2.[9] Here, since $[Type^{m_1}_\alpha(h_1)]^{m_2}_\alpha$ and $[Type^{m_2}_\alpha(h_3)]^{m_1}_\alpha$ are both defined, this model satisfies the weaker (C2), but since $Type^{m_1}_\alpha = \{\tau_1, \tau_2\}$ while $Type^{m_2}_\alpha = \{\tau_1\}$, it fails to satisfy the stronger (C1).

Even though the constraint (C1) is stronger than necessary simply to guarantee that the new operator is well-defined, it is a very natural constraint, which can be interpreted as requiring that an agent knows which action types are available for execution.[10] This requirement can be reflected in the object language if we introduce, for each agent α and action type τ, the special proposition A^τ_α, carrying the intuitive meaning that the agent α executes the action type τ, and governed by the following evaluation rule.

Definition 6 (Evaluation rule: A^τ_α) Where m/h is an index from a labeled stit model $\mathcal{M}$,

- $\mathcal{M}, m/h \models A^\tau_\alpha$ if and only if $Type^m_\alpha(h) = \tau$.

It is then easy to verify that models satisfying the (C1) constraint validate the formula

$$\Diamond A^\tau_\alpha \supset \mathsf{K}_\alpha \Diamond A^\tau_\alpha,$$

according to which, if it is possible for an agent to execute an action of a certain type, then the agent knows that. But this formula is not validated by models satisfying only the weaker (C2) constraint, such as the model depicted in Figure 6, where $A^{\tau_2}_\alpha$ holds at the index m_1/h_2, so that $\Diamond A^{\tau_2}_\alpha$ holds at m_1/h_1 but $\mathsf{K}_\alpha \Diamond A^{\tau_2}_\alpha$ does not.

[9] Another convention for interpreting figures: in diagrams depicting labeled stit frames and models, the type of an action token is written in the rectangle indicating that token.

[10] A similar constraint is suggested as Hypothesis 3 in Herzig and Troquard [11], who note that the idea is found also in Schobbens [22].

Having addressed the complications introduced by the evaluation rule for our new *kstit* operator, we can now illustrate this operator by returning to our previous coin game examples, making explicit the action types that were already implicit in our informal descriptions of these games. We will suppose, then, that $Type = \{\tau_1, \tau_2\}$, where, intuitively, τ_1 is the action type of betting heads and τ_2 is the action type of betting tails.

In the first two coin games, depicted in Figures 3 and 4, the concrete actions K_3 and K_5 are tokens of the type betting heads, while K_4 and K_6 are tokens of the type betting tails. We therefore have $[\tau_1]_\alpha^{m_2} = K_3$ and $[\tau_1]_\alpha^{m_3} = K_5$, and $[\tau_2]_\alpha^{m_2} = K_4$ and $[\tau_2]_\alpha^{m_3} = K_6$. Let us focus on the index m_2/h_1, where $Choice_\alpha^{m_2}(h_1)$ is K_3 and $Type_\alpha^{m_2}(h_1)$ is τ_1—the agent is performing the action token K_3 by executing the action type τ_1—so that $[Type_\alpha^{m_2}(h_1)]_\alpha^{m_2}$ is again K_3, by the execution/label constraints. In the initial game, from Figure 3, the agent is taken to know whether she occupies m_2 or m_3—no index in which either of these moment participates is indistinguishable from any index in which the other participates. Because of this, and because $[Type_\alpha^{m_2}(h_1)]_\alpha^{m_2} \subseteq |A|^{m_2}$, it follows that $[\alpha\ kstit\colon A]$ holds at the index m_2/h_1—the agent α sees to it that A even in the epistemic sense. In the second game, from Figure 4, the agent cannot distinguish m_2 from m_3. There is therefore a moment, m_3, participating in the indices m_3/h_3 and m_3/h_4, both indistinguishable from m_2/h_1, at which the action type executed at m_2/h_1 fails to guarantee A—we do not, that is, have $[Type_\alpha^{m_2}(h_1)]_\alpha^{m_3} \subseteq |A|^{m_3}$. Because of this, $[\alpha\ kstit\colon A]$ does not hold at m_2/h_1. Even though the action token performed by α at that index guarantees the truth of A—the formula $[\alpha\ stit\colon A]$ holds—the action type executed by α is not one that she knows to guarantee the truth of A, and so α fails to see to it that A in the epistemic sense.

In the third coin game, from Figure 5, the concrete actions K_5 and K_7 are tokens of the type betting heads, while K_6 and K_8 are tokens of the type betting tails. We therefore have $[\tau_1]_\alpha^{m_2} = K_5$ and $[\tau_1]_\alpha^{m_3} = K_7$, and $[\tau_2]_\alpha^{m_2} = K_6$ and $[\tau_2]_\alpha^{m_3} = K_8$. Focusing once more on the index m_2/h_1, we see again that $Type_\alpha^{m_2}(h_1)$ is τ_1. Because both $[\tau_1]_\alpha^{m_2} \subseteq |A|^{m_2}$ and $[\tau_1]_\alpha^{m_3} \subseteq |A|^{m_3}$, we have $[Type_\alpha^{m_2}(h_1)]_\alpha^{m'} \subseteq |A|^{m'}$ for each moment m' participating in an index that is indistinguishable from m_2/h_1, so that $[\alpha\ kstit\colon A]$ holds at m_2/h_1. Betting heads is an action type that α knows to guarantee a win, even though she does not know whether it is the nickel or the dime that has been placed heads up.

Finally—and returning to our motivating problem—we can now analyze the epistemic notion of ability by adapting our previous recipe of combining ordinary impersonal possibility with a stit operator, but in this case, appealing to the new epistemic *kstit*, rather than the familiar causal *stit*. The resulting proposal is that the formula

$$\Diamond[\alpha\ kstit\colon A]$$

can be taken to represent the idea that the agent α has the ability, in the epistemic sense, to see to it that A. This proposal yields the desired results in our previous examples: the formula is settled true at the moment m_2, for instance, in the games depicted in Figures 3 and 5, telling us in both cases that the agent has the ability to win in the epistemic sense, but settled false in the game from Figure 4, telling us that, although the agent can win in the causal sense, she does not have the ability to win in the epistemic sense.

5 Discussion

5.1 Some logical points

The epistemic *kstit* operator is strictly stronger than the causal *stit*: the formula

$$[\alpha \; kstit\text{: } A] \supset [\alpha \; stit\text{: } A]$$

is valid and its converse fails. The verification of validity is straightforward, but illustrates the function of the execution/label constraints in the definition of labeled stit models, and so is provided here. Suppose, then, that $[\alpha \; kstit : \; A]$ holds at an index m/h, so that $[Type^m_\alpha(h)]^{m'}_\alpha \subseteq |A|^{m'}$ for all m'/h' such that $m'/h' \sim_\alpha m/h$. Because $m/h \sim_\alpha m/h$, it follows that $[Type^m_\alpha(h)]^m_\alpha \subseteq |A|^m$. And because $Type^m_\alpha(h)$ is $Label(Choice^m_\alpha(h))$, it follows from the execution/label constraints that $[Type^m_\alpha(h)]^m_\alpha$ is just $Choice^m_\alpha(h)$. We therefore have $Choice^m_\alpha(h) \subseteq |A|^m$, so that $[\alpha \; stit : \; A]$ holds at m/h. To see that the converse of this formula fails, consider, for example, our second coin game from Figure 4, where $[\alpha \; stit\text{: } A]$ holds at m_2/h_1 but $[\alpha \; kstit\text{: } A]$ fails, since $m_3/h_3 \sim_\alpha m_2/h_1$ but we do not have $[Type^{m_2}_\alpha(h_1)]^{m_3}_\alpha \subseteq |A|^{m_3}$. Although the action token performed by α at m_2/h_1 happens to guarantee the truth of A at that index, the action type executed by the agent is not one that the agent knows to guarantee the truth of A, since there are indistinguishable indices at which it does not.

Even though the new *kstit* operator is, in general, strictly stronger than the familiar *stit*, things are different if we limit attention to models satisfying the additional constraint

(C3) If $m/h \sim_\alpha m'/h'$, then $m = m'$,

according to which indistinguishable indices must participate in the same moment, so that,

intuitively, the agent knows which moment she occupies. In this case, the implication displayed above can be strengthened to the equivalence

$$[\alpha\ \mathit{kstit}\colon A] \equiv [\alpha\ \mathit{stit}\colon A],$$

collapsing the epistemic *kstit* into the causal *stit*. The new *kstit* operator can therefore be seen as a conservative generalization of the ordinary *stit* operator: there is no difference between them as long as the agent knows everything about the past, leading up to the present moment—but they can come apart if the agent has any uncertainty about the past, in which case *kstit* is stronger. Since we are explicitly concerned in this paper with situations in which the agent lacks knowledge of past events, and so about which current moment she happens to occupy, we do not impose the (C3) constraint.

Although the familiar *stit* is a *KT45* modal operator, the new *kstit* is a normal operator satisfying the T schema

$$[\alpha\ \mathit{kstit}\colon A] \supset A,$$

but failing to satisfy both

$$[\alpha\ \mathit{kstit}\colon A] \supset [\alpha\ \mathit{kstit}\colon [\alpha\ \mathit{kstit}\colon A]]$$

$$\neg[\alpha\ \mathit{kstit}\colon \neg A] \supset [\alpha\ \mathit{kstit}\colon \neg[\alpha\ \mathit{kstit}\colon \neg A]],$$

the *4* and *5* schemata, respectively. A countermodel illustrating the failure of the *4* schema is depicted in Figure 7. Here, we can see that $[\alpha\ \mathit{kstit}\colon A]$ holds at m_1/h_1, since m_1 and m_2 are the only moments participating in indices indistinguishable from m_1/h_1 and we have both $[Type_\alpha^{m_1}(h_1)]_\alpha^{m_1} \subseteq |A|^{m_1}$ and $[Type_\alpha^{m_1}(h_1)]_\alpha^{m_2} \subseteq |A|^{m_2}$. The statement $[\alpha\ \mathit{kstit}\colon [\alpha\ \mathit{kstit}\colon A]]$, however, does not hold at m_1/h_1, since, although $m_2/h_4 \sim_\alpha m_1/h_1$, we do

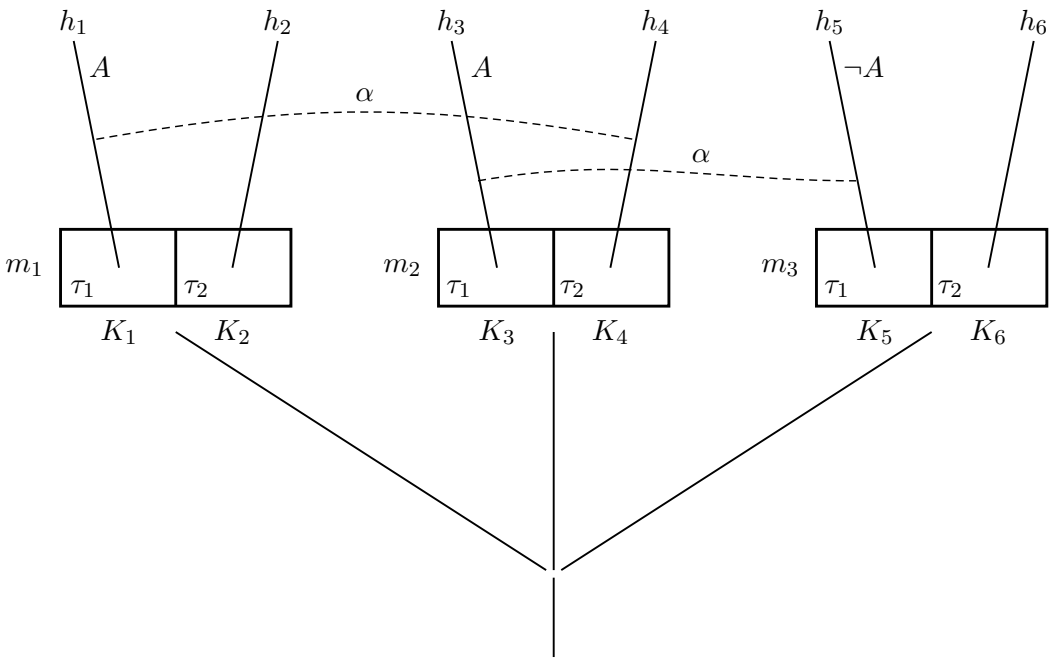

Figure 7: $[\alpha$ *kstit*: $A]$ without $[\alpha$ *kstit*: $[\alpha$ *kstit*: $A]]$

not have $[Type_\alpha^{m_1}(h_1)]_\alpha^{m_2} \subseteq |[\alpha\ kstit\colon A]|^{m_2}$. In particular, h_3 belongs to $[Type_\alpha^{m_1}(h_1)]_\alpha^{m_2}$, but h_3 does not belong to $|[\alpha\ kstit\colon A]|^{m_2}$, since $m_3/h_5 \sim_\alpha m_2/h_3$, but we do not have $[Type_\alpha^{m_2}(h_3)]_\alpha^{m_3} \subseteq |A|^{m_3}$.

As the model depicted in Figure 7 suggests, what makes it possible to falsify the *4* schema—and also the *5* schema—is the relative lack of constraint on the relation of epistemic indistinguishability, which plays a crucial role in the *kstit* evaluation rule. In this model, for example, an agent at the index m_1/h_2 is in the odd epistemic position of knowing that she is either at the moment m_1 executing the action type τ_1 or at the moment m_2 executing the action type τ_2, without knowing which.

We would like to propose that the *kstit* operator makes most sense, and is most useful, when the indistinguishability relation is systematized through the further constraint

(C4) If $m/h \sim_\alpha m'/h'$, then $m/h'' \sim_\alpha m'/h'''$ for all $h'' \in H^m$ and $h''' \in H^{m'}$,

according to which any indices involving moments that participate in indistinguishable indices must themselves be indistinguishable. The force of this constraint is that indistinguishability can be thought of as a relation, not just between indices, but between moments themselves, with $m \sim_\alpha m'$ defined to mean that $m/h \sim_\alpha m'/h'$ for all h from H^m and h' from $H^{m'}$. Our proposal, more precisely, is that the (C4) constraint provides an appropriate way of characterizing the knowledge of a deliberating agent who is aware that she occupies one of some definite set of moments, may not be sure which, and most important, has not yet decided which action type to execute in light of the available information. Of our earlier examples, the coin games depicted in Figures 3, 4, and 5 all satisfy (C4); only the artificial and uninterpreted examples from Figures 6 and 7 do not.[11]

Within the class of (C4) models, it turns out that *kstit* is a *KT45* modal operator, just like the ordinary *stit*, validating *4* and *5* as well as *T*. A proof of the *4* schema is provided here, simply to illustrate the role of the (C4) constraint. This proof relies on the observation that (*) $Type^{m'}_\alpha(h') = Type^m_\alpha(h)$ whenever h' belongs to $[Type^m_\alpha(h)]^{m'}_\alpha$, which itself follows from the execution/label constraints together with the fact that the *Label* function is one-one. To verify the *4* schema, then, suppose that $[\alpha\ kstit{:}\ A]$ holds at m/h, so that (**) $[Type^m_\alpha(h)]^{m'}_\alpha \subseteq |A|^{m'}$ for each m'/h' such that $m'/h' \sim_\alpha m/h$. In order to show that $[\alpha\ kstit{:}\ [\alpha\ kstit{:}\ A]]$ likewise holds at m/h, we must show that $[Type^m_\alpha(h)]^{m'}_\alpha \subseteq |[\alpha\ kstit{:}\ A]|^{m'}$ for each m'/h' such that $m'/h' \sim_\alpha m/h$. We therefore pick a particular m'/h' with $m'/h' \sim_\alpha m/h$, and supposing that h'' belongs to $[Type^m_\alpha(h)]^{m'}_\alpha$, argue that h'' belongs to $|[\alpha\ kstit{:}\ A]|^{m'}$ as well, or that

[11] The example from Figure 6 was introduced to distinguish the earlier (C1) and (C2) constraints, but in the presence of (C4), these two earlier constraints are equivalent.

(***) $[\alpha\ kstit\colon A]$ holds at the index m'/h''. To establish this latter claim, consider an index m''/h''' such that $m''/h''' \sim_\alpha m'/h''$. Because $m'/h' \sim_\alpha m/h$, the (C4) constraint along with transitivity of $\sim_\alpha$ now allows us to conclude that $m'/h'' \sim_\alpha m/h$, and then transitivity once again gives us $m''/h''' \sim_\alpha m/h$. From (**) we therefore have $[Type^m_\alpha(h)]^{m''}_\alpha \subseteq |A|^{m''}$, and then since h'' belongs to $[Type^m_\alpha(h)]^{m'}_\alpha$, we can conclude from (*) that $Type^m_\alpha(h) = Type^{m'}_\alpha(h'')$, so that $[Type^{m'}_\alpha(h'')]^{m'}_\alpha \subseteq |A|^{m''}$, which gives us (***).

Within this same class of (C4) models, we can see also that the statement

$$\mathsf{K}_\alpha[\alpha\ stit\colon A] \supset [\alpha\ kstit\colon A],$$

is valid and its converse invalid. Verification of this validity is straightforward, though it does rely on (C4); a counterexample to the converse can be found in our initial coin game from Figure 3, where $[\alpha\ kstit\colon A]$ holds at the index m_2/h_1 but $\mathsf{K}_\alpha[\alpha\ stit\colon A]$ does not, since $[\alpha\ stit\colon A]$ fails at the indistinguishable index m_2/h_2. This observation, taken together with our earlier observation that *kstit* is stronger than *stit*, shows that, in (C4) models, the statement $[\alpha\ kstit\colon A]$ lies properly between the statements $\mathsf{K}_\alpha[\alpha\ stit\colon A]$ and $[\alpha\ stit\colon A]$, strictly weaker than the first but strictly stronger than the second.

Just as we noted, earlier, the equivalence between $[\alpha\ kstit\colon A]$ and $[\alpha\ stit\colon A]$ under the constraint (C3), we can now note that the implication displayed just above can likewise be strengthened to the equivalence

$$\mathsf{K}_\alpha[\alpha\ stit\colon A] \equiv [\alpha\ kstit\colon A]$$

under the new constraint

(C5) If $m/h \sim_\alpha m'/h'$, then $Type^m_\alpha(h) = Type^{m'}_\alpha(h')$,

which requires that two indices can be indistinguishable for an agent only if the agent is executing the same action type at each. This constraint carries a good deal of initial plausibility, apparently capturing the attractive idea that an agent knows what she is doing; the idea can then seem to be reflected in the object language when we note that (C5) validates the formula

$$[\alpha \; \textit{kstit}\text{: } A] \supset \mathsf{K}_\alpha[\alpha \; \textit{kstit}\text{: } A]$$

according to which: whenever an agent executes an action type that she knows to guarantee the truth of A, she knows that she does this. In spite of its initial plausibility, however, we reject the (C5) constraint because, as the reader can easily verify, it is inconsistent with our earlier (C4) in any model in which an agent ever has more than one action to choose from.

But without (C5), what becomes of the attractive idea that an agent knows what she is doing? Here it is helpful to appeal to the distinction, familiar from economics, between *ex ante* knowledge and *ex interim* knowledge. Although these ideas have received various technical formulations, particularly within game theory, we understand them in the following sense, which we take to be consistent with their more technical treatments: an agent's *ex ante* knowledge is the information available to the agent without taking into account any actions she is currently executing, while the agent's *ex interim* knowledge is information that does take into account whatever actions the agent is currently executing, along with the effects of these actions.[12] Since, as we noted earlier, the (C4) constraint on the indistinguishability

[12]See Aumann and Dreze [3] for an authoritative discussion of the difference between *ex ante* and *ex interim* knowledge in game theory; a similar distinction is drawn in Stalnaker [24] between *passive* and *active* knowledge, where the former is defined as "knowledge based solely on observation, evidence and inference" and the latter as "knowledge based on decision."

relation $\sim_\alpha$ is appropriate for agents who have not yet decided which action type to execute, and since the K_α operator is keyed to this indistinguishability relation, it is natural to understand this particular knowledge operator as corresponding to *ex ante* knowledge. And once this interpretation of the operator is made explicit, it no longer appears that the implication displayed above, from $[\alpha\ kstit\colon A]$ to $\mathsf{K}_\alpha[\alpha\ kstit\colon A]$, should hold at all: even supposing the truth of $[\alpha\ kstit\colon A]$—that the agent α is executing an action type that she knows to guarantee the truth of A—it is hard to see why the agent should know this in the *ex ante* sense, without taking into account the action she is currently executing.

If the K_α operator is to be understood as carrying the *ex ante* sense of knowledge, is there another operator available to carry the *ex interim* sense? Yes. Our suggestion is that the *kstit* operator—expressing the idea that the agent executes an action type that she knows to guarantee a certain outcome—itself carries the *ex interim* sense of knowledge, since this operator mixes together knowledge and action, and indeed, can be thought of as refining the agent's *ex ante* knowledge by taking into account her current action and its effects. And if knowledge is interpreted in this sense, the *ex interim* sense carried by the *kstit* operator, the attractive idea that the agent knows what she is doing can now be rescued in the form of an implication from $[\alpha\ kstit\colon A]$ to $[\alpha\ kstit\colon [\alpha\ kstit\colon A]]$—from the idea that an agent executes an action type that she knows to guarantee the truth of A to the idea that the agent executes an action type that she knows to guarantee the truth of the fact that she executes an action type that she knows to guarantee the truth of A. This is simply the *4* schema, valid under the (C4) constraint and verified earlier.

Since *ex interim* knowledge is a refinement of *ex ante* knowledge, it is natural to expect

that the formula

$$\mathsf{K}_\alpha A \supset [\alpha\ \mathit{kstit}\colon A]$$

should be valid: if an agent knows that A in the *ex ante* sense, even without considering the action currently being executed, then the agent should still know A in the *ex interim* sense, when this additional information is taken into account. And this validity holds, in (C4) models. The converse implication, from *ex interim* to *ex ante* knowledge, fails, of course, since an agent might be able to conclude that A when information based on her current action is taken into account, without being able to conclude that A without that information. A formal counterexample, illustrating just this point, can be found in Figure 3, where $[\alpha\ \mathit{kstit}\colon A]$ holds at the index m_2/h_1 but $\mathsf{K}_\alpha A$ does not.

There is, however, one interesting implication from *ex interim* to *ex ante* knowledge, captured by the formula

$$\mathsf{K}_\alpha A \equiv \Box[\alpha\ \mathit{kstit}\colon A],$$

which is also valid in (C4) models. Here the left of right direction is simply a strengthening of the previous implication. The interesting direction is right to left, which tells us that, even though *ex interim* knowledge does not itself imply *ex ante* knowledge, it turns out that, if the agent has *ex interim* knowledge that A no matter which of her available actions she happens to execute, this entails that she must have *ex ante* knowledge that A. A verification is sketched here, since it makes interesting use of the (C1) constraint as well as (C4). Suppose, then, that (*) $\Box[\alpha\ \mathit{kstit}\colon A]$ holds at m/h. In order to see that $\mathsf{K}_\alpha A$ holds at m/h, we show that A holds at an arbitrary index m'/h' where $m'/h' \sim_\alpha m/h$. Since $m'/h' \sim_\alpha m/h$, it follows from (C1) that $\mathit{Type}^m_\alpha = \mathit{Type}^{m'}_\alpha$, from which we can conclude

that there is some history h'' through m such that the agent is executing the same action type at m/h'' and at m'/h'—or more formally, that (**) $Type^m_\alpha(h'') = Type^{m'}_\alpha(h')$. From (*), we know that $[\alpha\ kstit\colon A]$ holds at m/h'', and by (C4) and transitivity of $\sim_\alpha$ that $m'/h' \sim_\alpha m/h''$, from which it follows that $[Type^m_\alpha(h'')]^{m'}_\alpha \subseteq |A|^{m'}$. Together with (**), this yields $[Type^{m'}_\alpha(h')]^{m'}_\alpha \subseteq |A|^{m'}$. The execution/label constraints then allow us to conclude that $Choice^{m'}_\alpha(h') \subseteq |A|^{m'}$, from which it follows at once that A holds at m'/h'.

5.2 Connections

The problem motivating this paper—arriving at a satisfactory analysis of ability for agents with imperfect information—has been discussed extensively in the literature on multi-agent systems; the reader can consult, for example, the work on alternating-time temporal epistemic logic and concurrent game structures by Ågotnes [1], Jamroga and van der Hoek [15], Jamroga and Ågotnes [14], and Schobbens [22]. In spite of the technical differences between these frameworks and that of stit semantics, it is apparent that many of the proposals offered in that literature anticipate some of the ideas set out here. In particular, these proposals can all be seen as attempts at isolating the notion of a *uniform* action or strategy—an action or strategy that is, in some sense, constant across the various states that are indistinguishable by an agent, so that the action or strategy depends only on the agent's knowledge.[13] In the present paper, this notion of uniformity is captured through the introduction of action types, which can be thought of as executed uniformly across the set of indices indistinguishable to

[13]This use of the term "uniformity" in this context is, we believe, due to van Benthem [26], though of course the concept is familiar from game theory, where a strategy for an agent is defined as a function from that agent's information sets into moves, rather than from game states into moves.

an agent, leading to the performance of different action tokens depending on which particular indices the agent happens to occupy.

There is also, within the multi-agent systems literature, another line of research on action and ability in the face of imperfect information that is of special relevance to the present paper, both because of its intrinsic importance and also because it is carried out in the framework of stit semantics. This line of research was initiated by Herzig and Troquard in [11], further explored by Broersen in [5] and [6], and then developed in a number of fruitful directions by these three authors, working in various combinations and with various colleagues.[14] A detailed comparison is beyond the scope of this paper, but we do want to note that the idea of introducing action types into stit logic was already hinted at in the initial work of Herzig and Troquard, who describe action tokens as belonging to the same type if "*the way to produce them* or *the bodily movement part of the action* is the same."[15] Although Herzig and Troquard deserve credit for recognizing the importance of action types in stit semantics, and their overall way of thinking about types is very similar to ours, the appeal to types plays only a motivational role in their theory, without any real work to do in their formal machinery; action types are entirely absent in Broersen's work, and as far as we know, have not been explored any further within this line of research.[16]

[14]Our motivating coin examples from Section 3 of the present paper are similar in structure to Herzig and Troquard's light bulb examples first set out as Example 1 of their [11]; another similar situation is presented as Example 2.3 by Jamroga and van der Hoek [15].

[15]See Herzig and Troquard [11, Section 3.1, n. 3].

[16]An entirely different approach to action types within stit semantics, motivated by different concerns, is proposed by Xu [27]; a comparison between Xu's approach and that developed in the present paper would be interesting, but is not attempted here.

In the account set out here, by contrast, action types are present as first-class citizens within the semantic framework, and play a crucial role in the definition of our *kstit* operator. The reification of action types allows us to explore the relation between these types and other elements of the semantic framework, such as indistinguishability relations. We thus gain understanding at the cost of postulating some additional ontology—a tradeoff that is almost always worth making.[17]

6 Conclusion

Standard stit semantics has been criticized for focusing on agency at the expense of actions themselves—offering, as Lindström and Segerberg write, a "logic of action without actions." Our goal in this paper has been to extend this standard framework to the new framework of labeled stit semantics, providing an account of action types as well as action tokens, exploring the relations between these types and tokens, and introducing a new *kstit* operator corresponding to an epistemic sense of agency. We feel that this new framework of labeled stit semantics is faithful not only to the spirit of the standard framework, but also to its letter, since it collapses into this standard framework whenever the agent has perfect knowledge of the past, differing only in situations of uncertainty.

We have concentrated in this paper on motivation and preliminary definitions, but there

[17]It has recently come to our attention that Lorini, Longin, and Mayor [19] pursue a strategy similar to that explored here, introducing *names*, analogous to our labels, or action types, into a framework based on stit semantics. Although these authors are more interested in the study of a variety of epistemic operators, together with the role these operators play in attributions of responsibility, and although their work is carried out in a flat, atemporal setting, it is clear that our projects are closely related.

is much more substantial work to do. To begin with, the introduction of action types into the framework of stit semantics should allow for more explicit comparisons between this framework and those in which an understanding of agency is originally based on the idea of action type execution—the Davidsonian framework, of course, but especially the framework of dynamic logic. One of the advantages of standard stit semantics is the ease with which the theory it offers of individual agency generalizes to a theory of group agency, so it is natural to explore ways in which the account of individual epistemic agency presented here could be generalized to an account of group epistemic agency. Another advantage of standard stit semantics is the ease with which it can be adapted as the basis for a deontic logic, with ought statements defined in terms of a preference ranking on action tokens, so it is likewise natural to explore ways in which the framework of labeled stit semantics developed here could serve as the basis for a deontic logic in which epistemic ought statements are defined on the basis of a preference ranking of action types, relative to an agent's knowledge—and then further down the road, to an account of group epistemic oughts. These topics are currently under exploration.

Acknowledgments

We are especially grateful to Jan Broersen, Roberto Ciuni, Emiliano Lorini, Thomas Müller, Allard Tamminga, Nicholas Troquard, and Ming Xu for helping us find our way through this topic.

References

[1] Thomas Ågotnes. Action and knowledge in alternating-time temporal logic. *Synthese*, 149:377–409, 2006.

[2] Rajeev Alur, Thomas Henzinger, and Orna Kupferman. Alternating-time temporal logic. *Journal of the Association for Computing Machinery*, 49:672–713, 2002.

[3] Robert Aumann and Jacques-Henri Dreze. Rational expectations in games. *American Economic Review*, 98:72–86, 2008.

[4] Nuel Belnap, Michael Perloff, and Ming Xu. *Facing the Future: Agents and Choices in Our Indeterministic World*. Oxford University Press, 2001.

[5] Jan Broersen. A logical analysis of the interaction between 'obligation-to-do' and 'knowingly doing'. In Leendert van der Torre and Ron van der Meyden, editors, *Proceedings the Ninth International Workshop on Deontic Logic in Computer Science (DEON-08)*, volume 5076 of *Lecture Notes in Computer Science*, pages 140–154. Springer, 2008.

[6] Jan Broersen. Deontic epistemic *stit* logic distinguishing modes of mens rea. *Journal of Applied Logic*, 9(2):127 – 152, 2011.

[7] Jan Broersen, Andreas Herzig, and Nicolas Troquard. From coalition logic to STIT. In *Proceedings of the Third International Workshop on Logic and Communication in Multi-Agent Systems (LCMAS 2005)*, volume 157 of *Electronic Notes in Theoretical Computer Science*, pages 23–35. Elsevier, 2005.

[8] Jan Broersen, Andreas Herzig, and Nicolas Troquard. Embedding alternating-time temporal logic in strategic stit logic of agency. *Journal of Logic and Computation*, 16:559–578, 2006.

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

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

[11] Andreas Herzig and Nicolas Troquard. Knowing how to play: uniform choices in logics of agency. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS-06)*, pages 209–216. The Association for Computing Machinery Press, 2006.

[12] John Horty. *Agency and Deontic Logic*. Oxford University Press, 2001.

[13] John Horty and Nuel Belnap. The deliberative stit: a study of action, omission, ability, and obligation. *Journal of Philosophical Logic*, 24:583–644, 1995.

[14] Wojciech Jamroga and Thomas Ågotnes. Constructive knowledge: what agents can achieve under incomplete information. *Journal of Applied Non-Classical Logics*, 17:423–475, 2007.

[15] Wojciech Jamroga and Wiebe van der Hoek. Agents that know how to play. *Fundamenta Informaticae*, 63:185–219, 2004.

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

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

[18] Sten Lindström and Krister Segerberg. Modal logic and philosophy. In Patrick Blackburn, Johan van Benthem, and Frank Wolter, editors, *Handbook of Modal Logic*. Elsevier, 2007.

[19] Emiliano Lorini, Dominique Longin, and Eunate Mayor. A logical analysis of responsibility attribution: emotions, individuals, and collectives. *Journal of Logic and Computation*, 24:1313–1339, 2014.

[20] Marc Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12:149–166, 2002.

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

[22] Pierre-Yves Schobbens. Alternating-time logic with imperfect recall. *Electronic Notes in Theoretical Computer Science*, 85(2), 2004.

[23] Krister Segerberg. Getting started: beginnings in the logic of action. *Studia Logica*, 51:347–378, 1992.

[24] Robert Stalnaker. Extensive and strategic forms: games and models for games. *Research in Economics*, 53:293–319, 1999.

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

[26] Johan van Benthem. Games in dynamic epistemic logic. *Bulletin of Economic Research*, 53:219–249, 2001.

[27] Ming Xu. Combinations of stit and actions. *Journal of Logic, Language, and Information*, 19:485–503, 2010.