Epistemic Oughts in Stit Semantics

John Horty

*** Draft ***

Version of: April 5, 2018

Contents

1 Introduction

This paper explores some of the ways in which agentive, deontic, and epistemic concepts combine to yield ought statements—or simply, *oughts*—of different characters. Consider an example. Suppose I place a coin on the table, either heads up or tails up, though the coin is covered and you do not know which. And suppose you are then asked to bet whether the coin is heads up or tails up, with $10 to win if you bet correctly. If the coin is heads up but you bet tails, there is a sense in which we would naturally say that you ought to have made the other choice—at least, things would have turned out better for you if you had. But an ought statement like this does not involve any suggestion that you should be criticized for your actual choice. Nobody could blame you, in this situation, for betting incorrectly. By contrast, imagine that the coin is placed in such a way that you can see that it is heads up, but you bet tails anyway. Again we would say that you ought to have done otherwise, but this time it seems that you could legitimately be criticized for your choice.

These two scenarios have much in common: the coin is placed heads up but you bet tails, so that, in both cases, we would naturally say that you ought to have done otherwise. All that differs between the two scenarios is your knowledge—whether or not you know that the coin is heads up. Yet this difference is enough to influence the character of the resulting oughts, inviting criticism in one case but not the other.

The primary goal of the paper is to investigate agentive ought statements of the sort found in the second of these scenarios, where violation of the ought seems to invite criticism of the agent. Since an appeal to knowledge seems to play such an important role in the characterization of these statements, I refer to them as *epistemic oughts*.[1]

This investigation is not carried out in general, but in the particular setting of *stit semantics*, a framework for the analysis of agentive statements originating with a series of papers by

[1] This phrase has also been used to describe ought statements that are themselves concerned with epistemic matters, as when we say that an individual ought to know some fact, or ought not to believe some proposition as well as its negation. I do not discuss this use of the phrase here.

Nuel Belnap, Michael Perloff, and Ming Xu, culminating in their [6]. Stit semantics grows out of a modal tradition in the logic of action, going back to St. Anselm, but with more recent contributions by, among others, Alan Anderson, Brian Chellas, Frederic Fitch, Stig Kanger, Lars Lindahl, Ingmar Pörn, and Franz von Kutschera.[2] 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. Indeed, Sven Lindström and Krister Segerberg describe the work in this tradition as a "logic of action without actions," writing that "No author in the Anselm-Kanger-Chellas line up through Belnap ... has countenanced the existence of actions in logic: action talk, yes; ontology of action, no."[3]

This judgment is too strong. The framework of stit semantics does contain entities that can be regarded as action tokens—particular, concrete actions, each occurring at a single point in space and time. What Lindström and Segerberg mean to emphasize, however, is that, in contrast to the formal theories of action developed in the tradition of dynamic logic, which they favor, the standard framework of stit semantics makes no appeal to general, repeatable kinds of actions, or action types. There is not, for example, any general action type of "betting tails." There are only particular instances of betting tails—by particular individuals at particular moments in particular games—with nothing to group them together as actions of the same kind.

In recent work, Eric Pacuit and I [20] have argued that, in order to represent an epistemic sense of ability, it is helpful to enrich the standard framework of stit semantics with an explicit set of action types, in addition to the action tokens already present; the result is a new framework that we refer to as *labeled stit semantics*, where each action token is assigned a label, indicating the type of which it is a token. What the current paper shows is that an appeal to action types is likewise helpful in the analysis of epistemic oughts.

[2] A history of the subject, with references to the works of these writers and others, can be found in Segerberg [34], and at various points throughout Belnap, Perloff, and Xu [6].

[3] Lindström and Segerberg [24, p. 1199].

The paper is organized as follows: The next section summarizes the standard framework of stit semantics, leading to the definition of a standard stit operator representing individual agency. Section 3 then reviews an approach to agentive oughts in stit semantics set out in my previous [18], which relies on a preference ordering among the action tokens available to an individual. Section 4 explores the idea that epistemic oughts might by analyzed by combining this earlier approach with epistemic information in a particularly straightforward way, and points out problems with this initial proposal. This discussion motivates the introduction of action types in Section 5, which reviews the new framework of labeled stit semantics. Within this new framework, Section 6 suggests a new account of epistemic oughts that avoids the problems with the initial proposal; the account is similar in spirit to that set out earlier, but is based on an ordering of action types, rather than action tokens. Section 7 briefly explores two directions in which the account proposed here might be generalized: first to assessment sensitive, or relativistic, oughts, and then to conditional oughts.

2 Stit semantics

2.1 Branching time

Stit semantics is cast against the background of a theory of indeterministic time, first set out by A. N. Prior [28] and developed in more detail by Richmond Thomason [36], 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 number of different histories might pass through a single moment. 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 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 is 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 $\mathcal{M}, m/h \models A$ for each h from H^m; likewise A can be defined as *settled false* just in case $\mathcal{M}, m/h \models \neg A$ for each h from H^m. A formula is *determinant* at a moment m just in case it is either settled true or settled false at that moment, and *moment determinant* just in case it is determinant at every moment; a moment determinant statement can be thought of as one whose truth value depends only on the settled past, not on the open future.

Within the framework of branching time, there are several candidates available to play the role of the proposition expressed by a sentence. Two are mentioned here; some others will be introduced in Section 7. Perhaps the most natural proposal is that the *proposition expressed by the sentence* A in the model $\mathcal{M}$ should be identified with the entire set $|A|_{\mathcal{M}} = \{m/h : \mathcal{M}, m/h \models A\}$ of indices from that model in which A is true. But, although this global notion of a proposition may be natural, it is not especially helpful in the formal development of stit semantics. More useful is the moment relative notion of a proposition, based on the idea that the 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 then taken to represent worlds that are no longer accessible. On this view, the *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 along which that sentence is true.[4]

When context allows, we will omit reference to the background model in our notation, so that $m/h \models A$ can be taken to mean that A holds at the index m/h from some model that

[4]The relation between these two notions of a proposition is discussed in Section 2.1 of [18].

can be identified by context, or from an arbitrary model, and $|A|$ and $|A|^m$ can be taken to refer to the proposition expressed by A, or expressed by A at the moment m, from such a 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 the agent. In order to capture this idea, we must be able to speak of individual agents and 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. Now, what is it for one of these agents to act? Setting aside vagueness, probability, and many of the richer components of human action, stit semantics is based on the idea that acting, at a moment, is nothing more than constraining the course of future events to lie within some definite subset of the histories still available at that moment. These constraints are encoded through our second primitive: a function *Choice*, mapping each agent α and moment m to 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 selects 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.[5]

If K is a choice cell from $Choice^m_\alpha$, one of the equivalence classes belonging to 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

[5]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. The second 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. A thorough discussion of these requirements can be found at various points throughout Belnap, Perloff, and Xu [6] and in Chapter 2 of [18].

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.

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. Although stit frames and models can be very general, we simplify here in two ways. First, we suppose that, at any moment m, at most one agent faces a nontrivial choice—that is, that $Choice^m_\alpha \neq \{H^m\}$ for at most one agent α. Second, we suppose that any choice facing an agent involves only finitely many options—that $Choice^m_\alpha$ is always finite.

We can now introduce a *standard stit* operator—written, [... *stit*: ...]—allowing for statements of the form

$$[\alpha \textit{ stit}\text{: } A],$$

with the intuitive meaning that the agent 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 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 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:

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

- $\mathcal{M}, m/h \models [\alpha \textit{ stit}\text{: } A]$ if and only if $Choice^m_\alpha(h) \subseteq |A|^m_{\mathcal{M}}$.[6]

These various definitions are illustrated in Figure 1, which introduces the convention that a formula written next to some history emanating from a moment should be taken as true at

[6]Those familiar with stit logics will recognize the operator defined here as the "Chellas stit," first introduced into stit logics by Horty and Belnap [19], but drawing on ideas from Chellas [11].

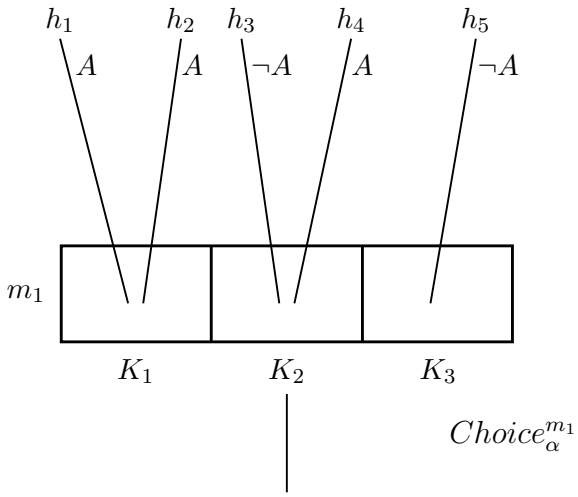

Figure 1: [α *stit*: A] true at m_1/h_1

that moment/history pair, so that A is true at m_1/h_1, m_1/h_2, and m_1/h_4, while $\neg A$ as true at m_1/h_3 and m_1/h_5, . As the diagram indicates, we have $Choice_\alpha^{m_1} = \{K_1, K_2, K_3\}$, with $K_1 = \{h_1, h_2\}$, $K_2 = \{h_3, h_4\}$, and $K_3 = \{h_5\}$. The statement [α *stit*: A] therefore holds at the index m_1/h_1, for example, since $Choice_\alpha^{m_1}(h_1) = K_1$ and $|A|^{m_1} = \{h_1, h_2, h_4\}$, so that $Choice_\alpha^{m_1}(h_1) \subseteq |A|^{m_1}$. But [$\alpha$ *stit*: A] does not hold at m_1/h_4, since $Choice_\alpha^{m_1}(h_4) = K_2$ and we do not have $Choice_\alpha^{m_1}(h_4) \subseteq |A|^{m_1}$. Even though the statement A itself happens to be true 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 Oughts in stit semantics

3.1 The Meinong/Chisholm analysis

This section presents a simplified version of the treatment set out in [18] of agentive oughts.

Typically in deontic logic, the ought operator $\bigcirc$ is interpreted against a background set of possibilities. Some nonempty group of these possibilities are classified as ideal, and a sentence of the form $\bigcirc A$—meaning, it ought to be the case that A—is then defined as true just in case A holds in each of these ideal possibilities. In the usual modal framework, this

approach leads to what is sometimes called *standard deontic logic*.[7]

This standard picture can be transposed into the current framework, and just slightly generalized, by identifying the possibilities at a moment m with H^m, the set of histories still available at m, and then introducing the function *Value* mapping each history h into a numerical value $Value(h)$, representing its overall worth, or desirability. A *deontic stit frame* can then be defined as a structure of the form

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

and a *deontic stit model* as a model based on a deontic stit frame. If we suppose, for simplicity, that there are only a finite number of values, rather than an ever-increasing series, the ideal possibilities available at a moment can be identified with those histories of greatest value, and an ought operator introduced by stipulating that $\bigcirc A$ ought holds at an index m/h whenever A holds at all the ideal histories through m.

Definition 4 (Evaluation rule: $\bigcirc A$) Where m/h is an index from a deontic stit model $\mathcal{M}$,

- $\mathcal{M}, m/h \models \bigcirc A$ if and only if $\mathcal{M}, m/h' \models A$ for each $h' \in H^m$ such that there is no $h'' \in H^m$ for which $Value(h') < Value(h'')$.

This definition leads to a normal modal logic, very similar to the standard theory, in which an ought statement is moment determinant, and in which the characteristic deontic principle that ought implies can holds in the form of the validity of $\bigcirc A \supset \Diamond A$.[8] It is important to emphasize, however, that this ought operator, like that from standard deontic logic, concerns only what ought to be the case, not what any particular agent ought to do about it. Supposing, for example, that all of the ideal histories through some moment are

[7] See Hilpinen and McNamara [17] for a survey that places standard deontic logic in a historical setting.

[8] Several systems of temporal deontic logic along the lines of that sketched here were introduced in the early 1980's by Åqvist and Hoepelman [3], Thomason [37], and van Eck [38]; the current presentation follows Thomason, generalizing only to allow histories of various values, rather than just two values.

histories in which it is warm and sunny tomorrow, the logic defined here would tell us that it ought to be warm and sunny tomorrow, without suggesting that any agent ought to see to it that it is warm and sunny, or that any agent could do this.

Still, even though this logic offers only an impersonal account of what ought to be the case, it is natural to suppose that personal, or agentive, ought statements could be arrived at by combining the impersonal ought defined here with a stit operator, representing agency. More exactly, it may seem natural to advance a proposal described in [18] as the *Meinong/Chisholm analysis*, after two prominent advocates, according to which the concept of what an agent ought to do can be identified with the concept of what it ought to be that the agent does.[9] In the current setting, the force of this proposal is that the statement $\bigcirc[\alpha \; \mathit{stit}\colon A]$, which carries the literal meaning that it ought to be the case that the agent α sees to it that A, can be taken as an analysis of the claim that α ought to see to it that A, or that seeing to it that A is something α ought to do.

This implementation of the Meinong/Chisholm analysis within stit semantics has much to recommend it.[10] Nevertheless, the analysis fails, and fails convincingly, as we can see by considering a pair of gambling examples.

Imagine, first of all, that an agent is faced with two options: gambling the sum of five dollars, or refraining from the gamble. If the agent gambles, we suppose there is a history in which she wins ten dollars and another in which she loses her original sum of five dollars; or she could refrain from the gamble, preserving her original sum. The situation is depicted in Figure 2, where α is the agent and m_1 is the moment at which she must choose either to gamble, by performing the action token K_1, or to refrain from gambling, by performing

[9] Chisholm's version of the proposal can be found in his [12]; Meinong's treatment is discussed by García [14]. Although Meinong and Chisholm are perhaps the most prominent advocates, the idea that what an agent ought to do can be represented through the combination of an impersonal deontic operator with an operator representing personal agency has also been suggested by a number of modern logicians, including Anderson [2], Kanger [22], and Hilpinen [16].

[10] Its virtues are discussed in Chapter 3 of [18].

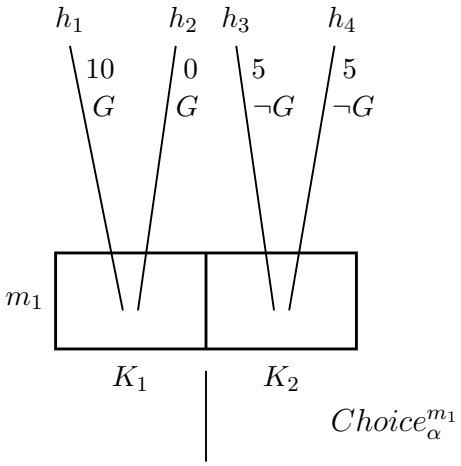

Figure 2: $\bigcirc[\alpha\ \mathit{stit}\colon G]$ settled true at m_1

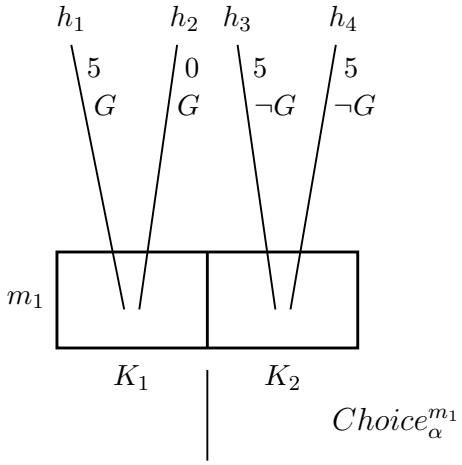

Figure 3: $\bigcirc[\alpha\ \mathit{stit}\colon \neg G]$ settled false at m_1

the action token K_2. Histories are assigned values corresponding to the resulting wealth of the agent, so that h_1, in which she gambles and wins, has the value 10, while h_2, in which she gambles and looses, has the value 0; the histories h_3 and h_4, in which she refrains from gambling, have the value 5. Finally, the statement letter G, true at m_1/h_1 and m_1/h_2, represents the proposition that the agent gambles.

In this situation, the statement $\bigcirc[\alpha \; \mathit{stit}\colon G]$ is settled true at the moment m_1, since $[\alpha \; \mathit{stit}\colon G]$ holds at m_1/h_1, where h_1 is the unique ideal history through m_1, the history of greatest value. The Meinong/Chisholm analysis of what an agent ought to do thus tells us that the agent ought to gamble. But this is a strange conclusion, since by gambling, the agent risks achieving an outcome of value 0, while she is able to guarantee an outcome of value 5 by refraining from the gamble. From an intuitive point of view, it seems to be impossible to say what the agent ought to do, and so we should reject any theory that makes a definite recommendation one way or the other.

A second example appears in Figure 3, which depicts a situation nearly identical to that from Figure 2, with agents, action tokens, and statement letters interpreted in the same way, but differing in values assigned to histories. Here, the gamble facing the agent is peculiar, since what she stands to earn if she gambles and wins is a mere five dollars, the very sum that she risks forfeiting if she gambles and loses. As a result, the history h_1 is now assigned the value 5, while h_2, h_3 and h_4 carry the same values as in Figure 2.

It seems clear that the agent in this situation should reject the gamble: why would she gamble, risking an outcome of value 0, simply for the chance of achieving an outcome no greater in value than one that she could guarantee by not gambling at all? A correct account of what the agent ought to do should therefore tell us that the agent ought not to gamble. But this is not the result generated by the Meinong/Chisholm analysis. Here, the statement $\bigcirc[\alpha \; \mathit{stit}\colon \neg G]$ is settled false at m_1, since $[\alpha \; \mathit{stit}\colon \neg G]$ does not hold in each ideal outcome—in particular, $[\alpha \; \mathit{stit}\colon \neg G]$ does not hold at m_1/h_1, where h_1 is ideal.

3.2 Agentive oughts

In response to these two examples, the idea explored in [18] was that a deontic operator representing what an agent ought to do at a moment could be defined, not directly in terms of the ordering on histories, but in terms of an ordering on the action tokens available to the agent at the moment—where this later ordering is itself defined in terms of the ordering on histories. There are, of course, many ways to lift an ordering from histories to an ordering on action tokens, or sets of histories.[11] The particular deontic logic presented in that book relies on a dominance ordering among action tokens, which is especially easy to define here in light of our first simplifying assumption on stit models, that at most one agent faces a nontrivial choice.

Suppose that K and K' are action tokens available to an agent at the moment m, subsets of the set H^m of histories through m, and that each history belonging to K' is at least as valuable as each history belonging to K. In that case, the action token K' can be said to weakly dominate the action token K, since the performance of K' is sure to result in an outcome at least as valuable as any resulting from the performance of K. And if we suppose, further, that the K does not itself weakly dominate K', then K' can be said to strongly dominate K, since, not only is the performance of K' sure to result in an outcome at least as valuable as any resulting from the performance of K, it might result in an outcome that is more valuable. These ideas are captured in the following definition:

Definition 5 (Dominance orderings on action tokens; $\leq, <$) Let α be an agent and m a moment from a deontic stit frame, and let K and K' belong to $Choice^m_\alpha$. Then $K \leq K'$ (K' weakly dominates K) if and only if $\mathit{Value}(h) \leq \mathit{Value}(h')$ for each $h \in K$ and $h' \in K'$; and $K < K'$ (K' strongly dominates K) if and only if $K \leq K'$ and it is not the case that $K' \leq K$.

[11] See Barberà, Bossert, and Pattanaik [4] for a general survey of methods for defining orderings on sets of objects in terms of orderings on the objects in those sets.

It is easy to see that both the weak and strong dominance orderings on action tokens are transitive, and that the strong ordering is, in addition, irreflexive.[12]

With this dominance ordering in place, we can now represent what an agent ought to do in a way that avoids the difficulties with the Meinong/Chisholm analysis. Syntactically, the idea is carried by an *agentive ought* operator—written, $\bigodot[\ldots\ \mathit{stit}\colon \ldots]$—allowing for construction of statements of the form

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

with the intuitive meaning that α ought to see to it that A.[13]

In order to describe the semantics of statements like these, we begin by defining the optimal action tokens available to an agent at a moment as those action tokens that are not strongly dominated by any others.

Definition 6 (Optimal action tokens; $K\text{-}Optimal^m_\alpha$) Where α is an agent and m a moment from a deontic stit frame, the optimal action tokens available to α at m are those belonging to the set

$$K\text{-}Optimal^m_\alpha = \{K \in Choice^m_\alpha : \text{there is no } K' \in Choice^m_\alpha \text{ such that } K < K'\}.$$

Because of our second simplifying assumption on stit models, that any choice involves only finitely many options—and because the strong dominance relation is transitive and irreflexive—the set of optimal action tokens available to an agent at a moment is guaranteed to be nonempty. The meaning of our agentive ought operator can therefore be defined very simply, through the stipulation that an agent ought to see to it that A just in case the truth of A is guaranteed by each optimal action tokens available to that agent.

[12]Neither ordering is linear: different actions available to an agent might be incomparable with respect even to weak dominance—a point illustrated by the example from Figure 2, where we have neither $K_1 \leq K_2$ nor $K_2 \leq K_1$.

[13]Note that this $\bigodot[\ldots\ \mathit{stit}\colon \ldots]$ operator, as well as later operators of the same pattern, is a single two-place operator, in which the symbol $\bigodot$ has no independent meaning.

Definition 7 (Evaluation rule: $\bigodot[\alpha\ stit\text{: } A]$) Where α is an agent and m/h is an index from a deontic stit model $\mathcal{M}$,

- $\mathcal{M}, m/h \models \bigodot[\alpha\ stit\text{: } A]$ if and only if $K \subseteq |A|^m_{\mathcal{M}}$ for each $K \in \textit{K-Optimal}^m_\alpha$.

Again, it can be shown that this definition leads to a normal modal logic, with statements of the form $\bigodot[\alpha\ stit\text{: } A]$ moment determinant, and with the characteristic deontic principle that ought implies can holding in the form of the validity of $\bigodot[\alpha\ stit\text{: } A] \supset \Diamond[\alpha\ stit\text{: } A]$. It is easy to see, also, that this new approach yields the correct results in the two gambling examples that led us to abandon the Meinong/Chisholm analysis. In the first example, from Figure 2, we wanted to avoid the conclusion that the agent α ought to gamble, and we can now do so, since both action tokens available to the agent are optimal, but they do not both entail gambling: more exactly, we have $\textit{K-Optimal}^{m_1}_\alpha = \{K_1, K_2\}$, but while $K_2 \subseteq |G|^{m_1}$, we do not have $K_2 \subseteq |G|^{m_1}$. As a result, the evaluation rule tells us that $\bigodot[\alpha\ stit\text{: } G]$ is settled false at the moment m_1. In the second example, from Figure 3, we wanted to reach the conclusion that the agent ought not to gamble, which we now do, since the only optimal action token available entails that the agent does not gamble: $\textit{K-Optimal}^{m_1}_\alpha = \{K_2\}$ and $K_2 \subseteq |\neg G|^{m_1}$, so that $\bigodot[\alpha\ stit\text{: } \neg G]$ is settled true at m_1.

4 Knowledge and oughts

4.1 An initial proposal

The logic just sketched, built around an ordering on action tokens, provides an account of what an agent ought to do that improves on the Meinong/Chisholm analysis. But the logic is less helpful in situations in which our evaluation of oughts is influenced by epistemic considerations.

To see this, we first incorporate epistemic information into the framework of stit semantics by adapting techniques that are, by now, standard in logic and game theory: we posit, for each agent α, an equivalence relation $\sim_\alpha$ among the moments from a stit frame, where

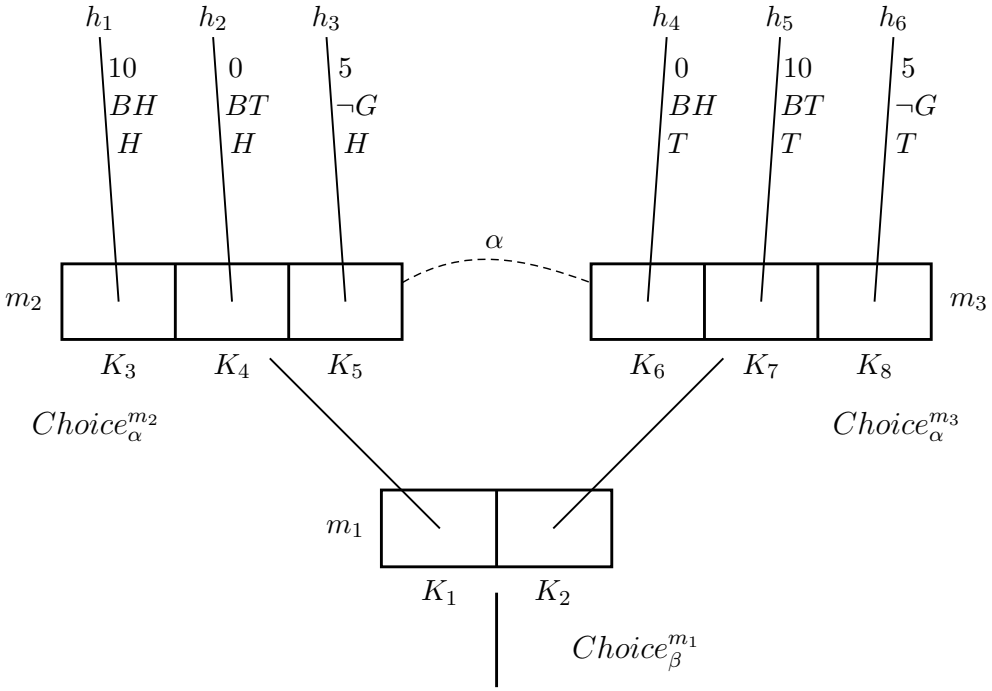

Figure 4: $\odot[\alpha \textit{ stit}\text{: } BH]$ settled true at m_2

$m \sim_\alpha m'$ is taken to mean that the moments m and m' are epistemically indistinguishable for α, or that nothing α knows allows her to distinguish m from m'.[14] Our previous deontic stit frames can then be supplemented with the additional component $\{\sim_\alpha\}_{\alpha \in Agent}$, a set containing indistinguishability relations for the various agents from $Agent$, leading to frames of the form

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

which are both deontic and *epistemic*; models can be defined, as usual, through the addition of a valuation function.

[14] Although we will work with this treatment of indistinguishability throughout the paper, it is worth noting that indistinguishability is typically cast as a relation between indices of evaluation, which, in the case of stit semantics, are moment/history pairs, rather than moments. The decision to treat indistinguishability as a relation between moments is a simplification, equivalent to understanding indistinguishability as a relation between moment/history pairs, in the more standard way, but subject to the (C4) constraint from Horty and Pacuit [20].

In this epistemic setting, let us now consider a situation very similar to the initial example from this paper, where I first place a coin on the table in such a way that you cannot see whether it is heads up or tails up, and you then bet on heads or tails. In the new situation, however, you are faced with a true gamble, not just an innocent bet: you must risk five dollars for the opportunity to bet on heads or tails, with ten dollars to win if you bet correctly and your original sum of five dollars to lose if you bet incorrectly; or you can choose not to gamble, preserving your original sum.

This new situation is depicted in Figure 4. Here, α represents you, β represents me, and m_1 is the moment at which I place the coin on the table, either heads up by performing the action token K_1, or tails up by performing K_2. Next, you must choose whether to bet heads, bet tails, or refrain from gambling. This action occurs at one of the later moments m_2 or m_3 in the branching time structure, depending on my initial choice at m_1. If I have placed the coin heads up, your choice occurs at the moment m_2, where you can bet heads by performing K_3, tails by performing K_4, or refrain from gambling by performing K_5. If I have placed the coin tails up, then your choice occurs at m_3, where you can bet heads by performing K_6, tails by performing K_7, or refrain from gambling by performing K_8. Of course, since you do not know, at the time of your choice, whether I have placed the coin heads up or tails up, the moments m_2 and m_3 are indistinguishable for you. We thus have $m_2 \sim_\alpha m_3$, indicated by an α-arc these two moments in the diagram.[15] The histories h_1 and h_5, in which you bet correctly, have the value 10, while h_2 and h_4, in which you bet incorrectly, have the value 0; the histories h_3 and h_6, in which you refrain from gambling, have the value 5.

Finally, the statements H and T stand for the respective propositions that I placed the coin heads up, or tails up; the first holds at the indices m_2/h_1, m_2/h_2, and m_2/h_3, while the second holds at m_3/h_4, m_3/h_5, and m_3/h_6. The statements BH and BT stand for the respective propositions that you bet heads or tails; the first holds at m_2/h_1 and m_3/h_4,

[15] Since indistinguishability is an equivalence relation, the actual indistinguishability relation at work in any particular stit frame is the closure of the relation explicitly depicted in the diagram of that frame under reflexivity, transitivity, and symmetry.

while the second holds at m_2/h_2 and m_3/h_5. The statement G, equivalent to the disjunction $BH \vee BT$, stands for the proposition that you gamble; this statement is true at any index where either BH or BT is true, and false at the indices m_2/h_3 and m_3/h_6, where you refrain from gambling.

Now suppose I placed the coin heads up, so that, at the time of your choice, you occupy the moment m_2. What ought you to do? According to the theory summarized in the previous section, the answer is unequivocal. Since K_3, the unique optimal action available to you at m_2, guarantees that you bet heads, you ought to bet heads: $\textit{K-Optimal}_{\alpha}^{m_2} = \{K_3\}$ and $K_3 \subseteq |BH|^{m_2}$, so that $\bigodot[\alpha \textit{ stit: } BH]$ is settled true at m_2. And indeed, as we noted earlier, there does seem to be a sense in which it is right to say, in this situation, that you ought to bet heads, since betting heads will result in an outcome of value 10, the greatest value available. But again, an ought statement like this, understood in this way, is not an epistemic ought—there is no suggestion that you should be criticized if you violate the ought. No one could blame you if you failed to bet heads, since you did not know that the coin had been placed heads up.

How, then, can we represent epistemic oughts, agentive ought statements that seem to be sensitive to the agent's knowledge, inviting criticism of the agent when violated? How do epistemic and deontic concepts interact in oughts like this? One very natural reaction to the situation just presented is to imagine that, although it may in fact be the case that you ought to bet heads, the reason you would not be criticized for failing to do so is simply that you do not know this—you do not know that you ought to bet heads. This reaction suggests an initial proposal according which criticism is tied, not to violations of what an agent in fact ought to do, but only to violations of what an agent knows she ought to do.

In order to capture this proposal formally, we must be able to speak explicitly of what an agent knows, or does not know. We therefore introduce, for each agent α, an operator K_α representing that agent's knowledge, so that a statement of the form $\mathsf{K}_\alpha A$ means that α knows that A. This knowledge operator is defined here in a standard fashion, adapted only

slightly to fit the framework of branching time, through the stipulation that an agent knows that A at an index m/h whenever A holds at every index m'/h' based on a moment m' that the agent cannot distinguish from m.

Definition 8 (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 $m' \sim_\alpha m$ and $h' \in H^{m'}$.

Once this knowledge operator has been introduced, our initial proposal can be set out as follows: the ought statements that matter in terms of criticism are not statements of the form $\bigodot[\alpha \; \mathit{stit}\colon A]$, describing what the agent in fact ought to do, whether she knows it or not, but statements of the form

$$\mathsf{K}_\alpha \bigodot[\alpha \; \mathit{stit}\colon A],$$

describing what the agent knows she ought to do. As we have seen, the statement $\bigodot[\alpha \; \mathit{stit}\colon A]$ is settled true at a moment m just in case each optimal action token available to the agent α at m guarantees the truth of A—just in case, that is, $K \subseteq |A|^m$ for each $K \in \mathit{K\text{-}Optimal}^m_\alpha$. By contrast, the statement $\mathsf{K}_\alpha \bigodot[\alpha \; \mathit{stit}\colon A]$ is settled true just in case each optimal action token available to α guarantees the truth of A, not just at m, but at every moment indistinguishable from m—just in case, for every moment m' such that $m' \sim_\alpha m$, we have $K \subseteq |A|^{m'}$ for each $K \in \mathit{K\text{-}Optimal}^{m'}_\alpha$.

This initial proposal provides us with a formal solution to the problem raised by the situation from Figure 4. What we needed to understand is why, when situated at m_2, you would not be criticized for failing to bet heads, even though you ought to bet heads—even though, that is, the statement $\bigodot[\alpha \; \mathit{stit}\colon BH]$ holds. And the answer provided by the initial proposal is that criticism is not warranted because you do not know that you ought to bet heads—the statement $\mathsf{K}_\alpha \bigodot[\alpha \; \mathit{stit}\colon BH]$ fails. We can verify this failure by noting that not every optimal action token available to you at every moment indistinguishable from m_2

guarantees that you bet heads. Since you do not know whether I have placed the coin heads up or tails up, the moments you cannot distinguish from m_2 are m_2 itself and m_3, with the optimal action tokens available to you at these moments calculated as: $\textit{K-Optimal}_{\alpha}^{m_2} = \{K_3\}$ and $\textit{K-Optimal}_{\alpha}^{m_3} = \{K_7\}$. And while your unique optimal action at m_2 guarantees that you bet heads, your unique optimal action at m_3 does not: while $K_3 \subseteq |BH|^{m_2}$, we do not have $K_7 \subseteq |BH|^{m_3}$.

4.2 Problems with the initial proposal

The initial proposal—that criticism is tied to knowledge of oughts, rather than oughts themselves—has a good deal of intuitive appeal, and seems to offer a satisfying solution to the initial problem raised by our example. But the proposal fails. It does not provide an adequate account of the way in which epistemic and deontic concepts interact to yield oughts whose violations invite criticism, as we can see by considering three further problems.

To understand the first of these problems, we need only look a bit more closely at the situation depicted in Figure 4, supposing again that you occupy m_2. As we have just seen, the optimal action tokens available to you at the moments indistinguishable from the moment you occupy are $\textit{K-Optimal}_{\alpha}^{m_2} = \{K_3\}$ and $\textit{K-Optimal}_{\alpha}^{m_3} = \{K_7\}$, with the result that the statement $\mathsf{K}_{\alpha}\bigodot[\alpha \textit{ stit: } BH]$ is settled false at m_2. You do not know that you ought to bet heads, since not all of these optimal action tokens guarantee that you bet heads. In the same way, we can see that the statement $\mathsf{K}_{\alpha}\bigodot[\alpha \textit{ stit: } BT]$ is settled false as well. You do not know that you ought to bet tails, since not all of these optimal action tokens guarantee that you bet tails: while $K_7 \subseteq |BT|^{m_3}$, we do not have $K_3 \subseteq |BT|^{m_2}$. But now, recall the statement letter G, equivalent to $BH \vee BT$, representing the proposition that you gamble. It turns out that the statement $\mathsf{K}_{\alpha}\bigodot[\alpha \textit{ stit: } G]$ is settled true at m_2. On the current analysis, that is, you do know that you ought to gamble, since each of the optimal action tokens available to you at any moment indistinguishable from m_2 guarantees that you either bet heads or bet tails, and both betting heads and betting tails are ways of gambling: $K_3 \subseteq |G|^{m_2}$ and

$K_7 \subseteq |G|^{m_3}$.

This is already bad enough, since it does not seem right to say, in this situation, that you know you ought to gamble—the fact that the statement $\mathsf{K}_\alpha \bigodot [\alpha \; \textit{stit}\text{: } G]$ is true suggests that it does not even properly capture the intuitive idea that you know you ought to gamble. And things get even worse when we recall that, according to our initial proposal, criticism is tied to violation of the statement $\mathsf{K}_\alpha \bigodot [\alpha \; \textit{stit}\text{: } G]$. Even though this statement is true in the current situation, it does not seem that you could be criticized if you choose not to gamble, since, by gambling, you risk an outcome of value 0 while you could guarantee an outcome of value 5 by refraining from the gamble.

The reader will note the similarity between this problem with the initial proposal and our first objection to the Meinong/Chisholm analysis, centered around the example from Figure 2. Likewise, the second problem with the initial proposal is modeled after our second objection to the Meinong/Chisholm analysis, centered around the peculiar gamble depicted in Figure 3, where the value an agent stands to gain if she gambles and wins is no greater than the value she must risk to gamble at all.

This second problem is illustrated in Figure 5, which depicts a situation nearly identical to that from Figure 4, with agents, action tokens, and statement letters interpreted in the same way, differing only in the values assigned to histories. Here again, the gamble you face is peculiar: all you stand to gain if you gamble and win is five dollars, the sum you must risk to gamble at all. As a result, the histories h_2 and h_4, in which you bet incorrectly, continue to carry the value 0, and the histories h_3 and h_6, in which you refrain from gambling, continue to carry the value 5. But in this case, the histories h_1 and h_5, in which you bet correctly, carry the value 5 as well.

Again, suppose that I have placed the coin heads up, so that you occupy the moment m_2, though you do not know this, since you cannot distinguish m_2 from m_3. The optimal action tokens available to you at these indistinguishable moments can be calculated as: $\textit{K-Optimal}_\alpha^{m_2} = \{K_3, K_5\}$ and $\textit{K-Optimal}_\alpha^{m_3} = \{K_7, K_8\}$. And even though some of these

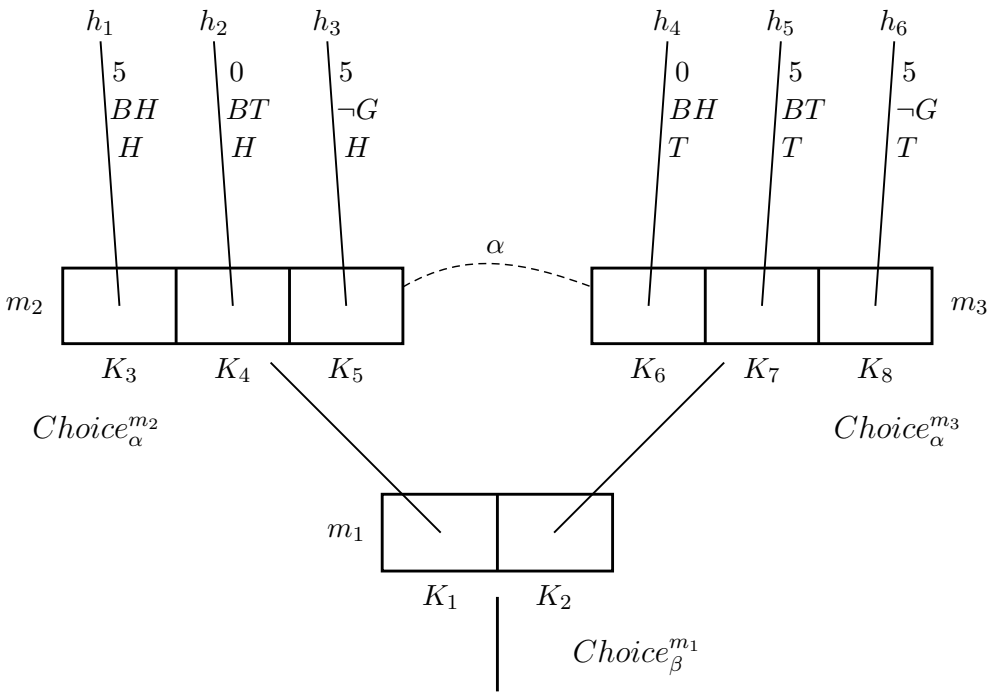

Figure 5: $\mathsf{K}_\alpha\bigodot[\alpha\ \mathit{stit}\colon \neg G]$ settled false at m_2

optimal action tokens guarantee that you refrain from gambling, others do not: while $K_5 \subseteq |\neg G|^{m_2}$ and $K_8 \subseteq |\neg G|^{m_3}$, we have neither $K_3 \subseteq |\neg G|^{m_2}$ nor $K_7 \subseteq |\neg G|^{m_3}$. It therefore follows that the statement $\mathsf{K}_\alpha\bigodot[\alpha\ \mathit{stit}\colon \neg G]$ is settled false at m_2—you do not know that you ought not to gamble.

As before, this result is objectionable from the start, since it seems natural to conclude that you do know that you ought not to gamble, or at least that an ideal reasoner would know that. And again, it becomes even more problematic when we consider that, according to the initial proposal, criticism is tied to a violation of the statement $\mathsf{K}_\alpha\bigodot[\alpha\ \mathit{stit}\colon \neg G]$. Since this statement is false, you would not run afoul of what it requires by gambling, so that, according to the initial proposal, criticism would not be appropriate. But it does seem, in this situation, that criticism of gambling is appropriate: why would you gamble, hoping to achieve an outcome of value of 5 but risking an outcome of value 0, when you could guarantee an outcome of value 5 by refraining from the gamble?

To understand the third problem with the initial proposal, we return to the situation depicted in Figure 4, once more supposing that you occupy m_2. Let us now take the new statement letter W, equivalent to the formula $(BH \wedge H) \vee (BT \wedge T)$, to represent the proposition that you win, or bet correctly; this statement is true at the indices m_2/h_1 and m_3/h_5 and nowhere else. As we have seen, the optimal action tokens available to you at the moments you cannot distinguish from m_2 are the members of $\textit{K-Optimal}_{\alpha}^{m_2} = \{K_3\}$ and $\textit{K-Optimal}_{\alpha}^{m_3} = \{K_7\}$, and both of these optimal actions guarantee that you win: $K_3 \subseteq |W|^{m_2}$ and $K_7 \subseteq |W|^{m_3}$. As a result, the statement $\mathsf{K}_{\alpha}\bigodot[\alpha \textit{ stit: } W]$ is settled true at m_2—you know that you ought to win.

The truth of this statement points to a different kind of problem for the initial proposal, according to which criticism is tied to violation of statements of this form. In the previous two cases, it seemed that criticism would be inappropriate because the initial proposal yielded the wrong results—telling us in the first case, incorrectly, that you know you ought to gamble, and then in the second case, failing to tell us that you know you ought not to gamble. In this third case, it is not so much that the results of the proposal are evidently incorrect. Perhaps you do know that you ought to win—perhaps this statement reflects some sort of conceptual truth about gambling. The problem in this case is that, all the same, it does not seem that you could legitimately be criticized for failing to win. Why not? Well, it seems safe to suppose that you can legitimately be criticized for failing to do something only if it is something you are able to do. This idea is often captured with the slogan that ought implies can, or in the presence of agency, that ought implies ability. But in the current example, winning simply does not seem to be something that lies within your ability—it does not seem to be an outcome that you are able to guarantee. And if you are not able to win, you cannot legitimately be criticized for failing to do so.

5 Labeled stit semantics

5.1 Action types and ability

The argument just offered—that you cannot be criticized for failing to win because you are not able to win—may seem to be too quick. One might object that, at the moment m_2 in the situation from Figure 4, you are, in fact, able to win. You could perform the action token K_3, in which case you would win—indeed, the statement $\Diamond[\alpha \; \mathit{stit}\colon W]$, which is taken in [18] to represent the proposition that you have the ability to win, is settled true at m_2. But this objection turns on an ambiguity. There is a sense, captured by the truth of $\Diamond[\alpha \; \mathit{stit}\colon W]$, in which you do have the ability to win—Pacuit and I refer to this in [20] as the *causal* sense of ability. But since you do not know whether the coin is heads up or tails up, there is another sense of ability—the *epistemic* sense—in which you do not have this ability, and it is this latter sense that seems to be crucial for assessing the legitimacy of criticism.

In trying to understand this epistemic sense of ability, we run up against a limitation of the standard stit framework: its restriction to action tokens. From an intuitive standpoint, what you face in the situation from Figure 4 are three options: betting heads, betting tails, or refraining from gambling. These three options must be thought of as action types, rather than tokens, since their execution at different moments results in the performance of different action tokens. If you execute the action type of betting heads, for example, you can hope you occupy the moment m_2, in which case the action token you perform is K_3 and you win, but you might occupy m_3, in which case the action token you perform is K_6 and you do not win.

The analysis of the epistemic sense of ability set out in [20] makes crucial use of action types and their interaction with the agent's knowledge. According to this analysis, you have the ability to win, in the epistemic sense, just in case there is an action type available to you whose execution you know will guarantee that you win. The reason you do not have this ability in the situation under consideration, then, is simply that, because you do not know

whether the coin is heads up or tails up, and so whether you occupy m_2 or m_3, you do not know which of the relevant action types—betting heads or betting tails—will guarantee that you win.

The remainder of this section summarizes the approach developed in that paper, extending the framework of stit semantics to include types as well as tokens, defining a new epistemic stit operator that draws on these action types, and also a new formula to capture the epistemic sense of ability.[16]

5.2 The *kstit* operator

We begin by explicitly postulating a set $Type = \{\tau_1, \tau_2, \ldots, \tau_n\}$ of action types—general kinds of action, as opposed to the concrete action tokens already present in stit logics. We assume here, for simplicity only, that there are a finite number of action types and that all action types are primitive.[17] In contrast to action tokens, action types are repeatable. A robot 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 semantics, 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 Figure 4. It is hard to see, in this situation, how you could actually choose to

[16] Although we focus here only on the approach of [20], it must be noted that this approach continues a line of research initiated by several computer scientists on epistemic concepts of action and ability in stit semantics; see, for example, Broersen [7], Herzig and Troquard [15], and Lorini, Longin, and Mayor [25]. This line of research was itself motivated by the problem of arriving at a satisfactory analysis of ability for agents with imperfect information in game theory and multi-agent systems; see, for example, Ågotnes [1], Jamroga and van der Hoek [21], and Schobbens [33].

[17] These assumptions could be relaxed in a more general setting, perhaps allowing for an infinite number of complex action types specified by a compositional action description language.

perform the action token K_3, for example, since that action token is not available to you unless you occupy the moment m_2, and you do not know whether you occupy m_2 or m_3. All you can do is execute the action type of betting heads, which will then result in the performance of the token K_3 if you are at m_2 and K_6 if you are 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 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 the type of action under which this particular token falls. 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 whose type is $Label(K)$, then the execution of that action type by α at m results in the performance of K itself; the second requires that, if τ is an action type whose execution by α at m results in the performance of the 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 deontic stit frame* as a structure of the form

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

containing all the components introduced earlier, as well as $Type$, $[\ ]$, and $Label$ as specified above; a *labeled deontic stit model* results when such a frame is supplemented with a valuation.

And we can then introduce a new *epistemic stit* operator—written, $[\ldots\ kstit{:}\ \ldots]$—allowing for statements such as

$$[\alpha\ kstit{:}\ A].$$

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 $[\alpha\ stit{:}\ A]$ was taken to mean that α performs an action token guaranteeing the truth of A, what $[\alpha\ kstit{:}\ A]$ means, somewhat roughly, 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' that is indistinguishable for α from m. The action type executed by α at the index m/h is $Type^m_\alpha(h)$, as we have seen, and so the execution of this action type by α at another moment m' is $[Type^m_\alpha(h)]^{m'}_\alpha$. The evaluation rule for our new operator, therefore, is as follows.

Definition 9 (Evaluation rule: $[\alpha\ kstit\text{: } A]$) Where α is an agent and m/h an index from a labeled deontic stit model $\mathcal{M}$,

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

This rule introduces a complication, which we can see by noting that it 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 being indistinguishable for the agent. 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 available for execution also at m'. We therefore stipulate that labeled stit frames must satisfy the *type/indistinguishability* constraint

If $m' \sim_\alpha m$, then $Type^{m'}_\alpha = Type^m_\alpha$,

according to which the same action types must be available for execution by an agent at any two moments that are indistinguishable for that agent; the intuitive force of this constraint is that an agent must know which action types are available for execution.[18]

The new *kstit* operator, representing an epistemic sense of agency, can be illustrated, and contrasted with the original *stit* operator, representing agency only in a causal sense, by returning to the situation from Figure 4. Ignoring the uninteresting actions available to me of placing the coin on the table either heads up or tails up, and considering only the interesting actions available to you, we take $Type = \{\tau_1, \tau_2, \tau_3\}$, where τ_1 is the action type of betting heads, τ_2 is the action type of betting tails, and τ_3 is the action type of refraining from gambling. As our informal description makes clear, the concrete actions K_3 and K_6 are tokens of the type betting heads, K_4 and K_7 are tokens of the type betting tails, and

[18]This constraint is the (C1) constraint from [20], where it is discussed in more detail, along with other options for constraining the relation between types and indistinguishability.

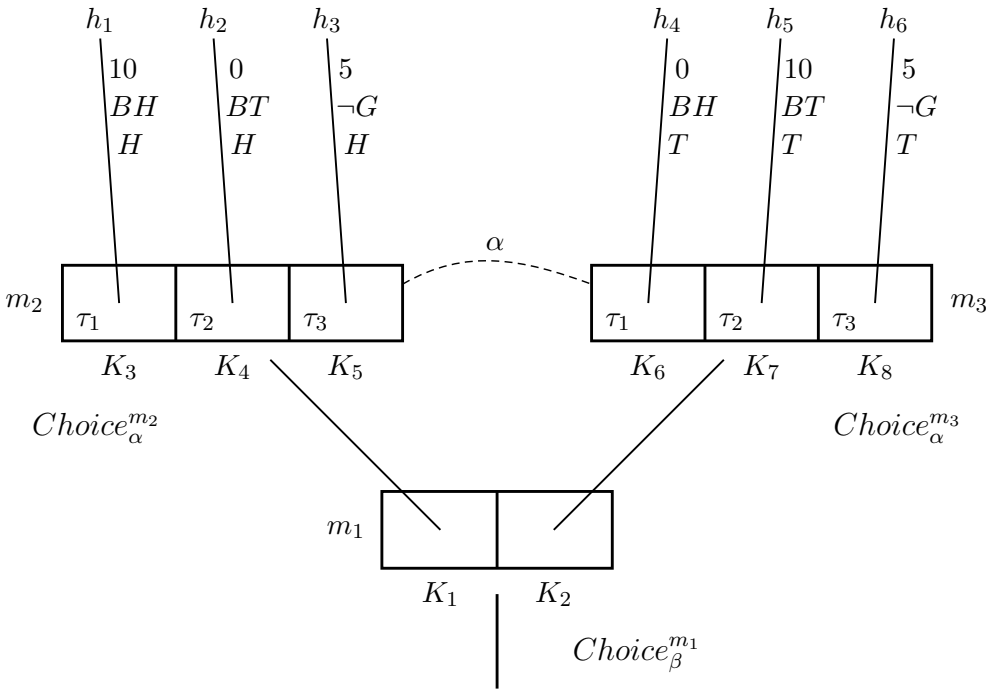

Figure 6: Action types

K_5 and K_8 are tokens of the type refraining from gambling. We therefore have $[\tau_1]_\alpha^{m_2} = K_3$ and $[\tau_1]_\alpha^{m_3} = K_6$, $[\tau_2]_\alpha^{m_2} = K_4$ and $[\tau_2]_\alpha^{m_3} = K_7$, and $[\tau_3]_\alpha^{m_2} = K_5$ and $[\tau_2]_\alpha^{m_3} = K_8$. This information appears in Figure 6, with the action types implicit in our informal description of the situation now displayed explicitly, in accord with the convention that the type of an action token is written inside the rectangle indicating that token.

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—you are performing the action token K_3 by executing the action type τ_1. Recall that W, equivalent to $(BH \wedge H) \vee (BT \wedge T)$, stands for the proposition that you win, and that G, equivalent to $BH \vee BT$, stands for the proposition that you gamble. Because $K_3 \subseteq |W|^{m_2}$, the statement $[\alpha\ \textit{stit}\colon W]$ is true at this index—you see to it that you win in the causal sense captured by the ordinary *stit* operator, since the action token you perform guarantees the truth of W. But the statement $[\alpha\ \textit{kstit}\colon W]$ is false—you do not see to it that you win in the epistemic sense captured by the *kstit* operator, since there are moments indistinguishable for you from the one you occupy at which τ_1, the action type you execute, results in the performance of an

action token that does not guarantee the truth of W. In particular, you cannot distinguish m_3 from m_2, and $[\tau_1]_\alpha^{m_3} = K_6$, as we have seen, but we do not have $K_6 \subseteq |W|^{m_3}$. On the other hand, the statement $[\alpha\ kstit\text{: } G]$ is true at m_2/h_1—you do see to it that you gamble even in the epistemic sense, since the action type you execute at this index results, at each moment indistinguishable for you from m_2, in the performance of an action token that guarantees the truth of G. The only moments indistinguishable for you from m_2 are m_3 and m_2 itself; again, $[\tau_1]_\alpha^{m_2} = K_3$ and $[\tau_1]_\alpha^{m_3} = K_6$, and we have both $K_3 \subseteq |G|^{m_2}$ and $K_6 \subseteq |G|^{m_3}$.

We refrain from discussing the logic of the *kstit* operator, referring the reader to [20] for details, except to offer two observations. First, the epistemic *kstit* is strictly stronger than the ordinary causal *stit*, but strictly weaker than a knowledge operator combined with the ordinary *stit*; more exactly, both the formulas

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

are valid in labeled stit models, and both converses fail. And second, if we assume that the agent always knows which moment she occupies, then the new *kstit* collapses into the ordinary *stit*; or more exactly, restricting attention to labeled stit models satisfying the *perfect information* constraint

If $m' \sim_\alpha m$, then $m' = m$,

the second of the above implications can be strengthened to the equivalence

$$[\alpha\ kstit\text{: } A] \equiv [\alpha\ stit\text{: } A].$$

The *kstit* operator can therefore be seen as a conservative extension of the ordinary *stit*. There is no difference between these two operators 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 past events, in which case the *kstit* operator is stronger.[19]

[19]The *perfect information* constraint at work here is the constraint (C3) from [20].

Finally, drawing on the new *kstit* operator, we can now analyze the epistemic notion of ability by taking statements of the form $\Diamond[\alpha\ kstit\text{: } A]$ to represent the idea that the agent α has the ability, in the epistemic sense, to see to it that A. Returning to our motivating example, the reader can verify that, although the statement $\Diamond[\alpha\ stit\text{: } W]$ is settled true at the moment m_2 from Figure 6, supporting the idea that you do have the ability to win in the causal sense, the statement $\Diamond[\alpha\ kstit\text{: } W]$ is settled false, since you do not have the ability to win in the epistemic sense.

6 Epistemic oughts

6.1 Ordering the action types

We now turn to our central topic: the definition of an epistemic ought operator based on a preference ordering of action types, rather than action tokens. The definition proceeds relative to the notion of an *information set bearing on an agent* α—or where clarity allows, simply an *information set*—defined as a nonempty set I of moments subject to the *type/information* constraint

If $m, m' \in I$, then $Type^m_\alpha = Type^{m'}_\alpha$.

A set of this kind, subject to this constraint, can be thought of as representing the information that the agent α occupies some moment belonging to the set, with the same set of action types available throughout. Of course, if an information set represents a body of information, we must ask exactly whose information this is supposed to be. In defining the epistemic oughts pertaining to an agent α at a moment m, we will concentrate on information sets of the form

$$I^m_\alpha = \{m' : m \sim_\alpha m'\},$$

representing the information available to the agent herself, at that moment.[20] Our initial definitions, however, will be developed in terms of an arbitrary information set, in order to

[20]Because labeled stit models are subject to the type/indistinguishability constraint set out earlier, in Section 5.2, it follows that I^m_α satisfies the type/information constraint.

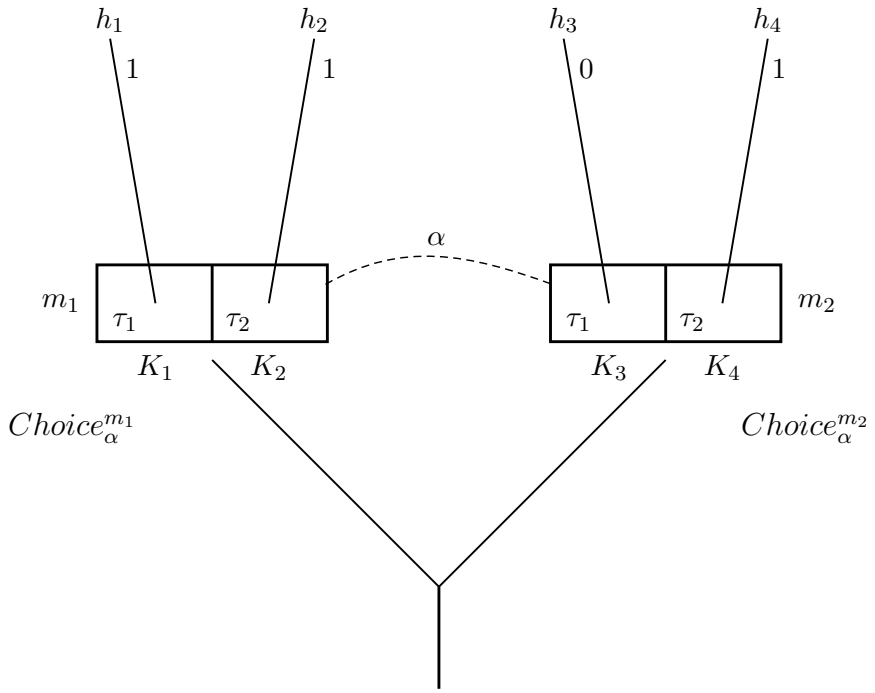

Figure 7: Ordering the action types

allow for later generalizations.

How should we define a preference ordering on the action types available to an agent at the moments from an information set, on the basis of the information only that the agent occupies one of those moments? We begin with an example, depicted in Figure 7, with $I = \{m_1, m_2\}$ as an information set bearing on the agent α. What this information set tells us is that the agent α, with the action types τ_1 and τ_2 available for execution, occupies either the moment m_1 or the moment m_2. If the agent executes the action type τ_1, the outcome will be the history h_1, with value 1, if she occupies the moment m_1, or the history h_3, with value 0, if she occupies the moment m_2; likewise, if the agent executes the action type τ_2, the outcome will be h_2 if she occupies m_1, or h_4 if she occupies m_2, both with value 1. Since the information set I does not specify whether the agent occupies m_1 or m_2, it does not allow us to determine whether the outcome of τ_1 will be h_1 or h_3, or whether the outcome of τ_2 will be h_2 or h_4. Nevertheless, it does seem that we can conclude on the basis of this information that τ_2 is a better action type for the agent to execute than τ_1,

since τ_2 guarantees an outcome of value 1 while τ_1 promises no better than 1 but allows the possibility of 0.

In light of this example, it may seem tempting to propose that two action types available to an agent at the moments throughout an information set should be ordered by comparing the entire set of outcomes that might, consistent with that information, issue from the execution of one of these action types with the entire set of outcomes that might issue from the execution of the other. To put this proposal precisely, we note first that the set $\{[\tau]^m_\alpha : m \in I\}$ contains all the action tokens that might result when the action type τ is executed by the agent α at some moment from I, so that the set

$$[\tau]^I_\alpha \;=\; \bigcup\{[\tau]^m_\alpha : m \in I\}$$

is the union of the histories belonging to these action tokens—that is, the entire set of outcomes that might issue from the execution of τ by α at some moment from I. Our proposal can then be captured through an ordering according to which, of two action types τ and τ' available to α throughout I, the action type τ' is at least as good as τ, on the basis of this information, just in case each history from $[\tau']^I_\alpha$ is at least as valuable as each history from $[\tau]^I_\alpha$, and τ' is better than τ just in case τ' is at least as good as τ and the reverse does not hold.

This proposal supports our intuition in the current example, from Figure 7. Here, the action type τ_2 is ordered as strictly better than τ_1 on the basis of the information set $I = \{m_1, m_2\}$, since each history from $[\tau_2]^I_\alpha = \{h_2, h_4\}$ is at least as valuable as each history from $[\tau_1]^I_\alpha = \{h_1, h_3\}$, and the reverse does not hold. The proposal does not always yield the right result, however, as we can see from another example, depicted in Figure 8.[21] This example is structurally similar to the previous case but with different values assigned to histories—the histories h_1, h_2, h_3, and h_4 are now assigned the respective values 4, 5, 9, and 10. In this new example, it no longer holds that each history from $[\tau_2]^I_\alpha = \{h_2, h_4\}$ is at least as valuable as each history from $[\tau_1]^I_\alpha = \{h_1, h_3\}$, since h_2 is less valuable than h_3, nor does it hold that

[21] For the moment, the reader should ignore the sentence letters in this diagram.

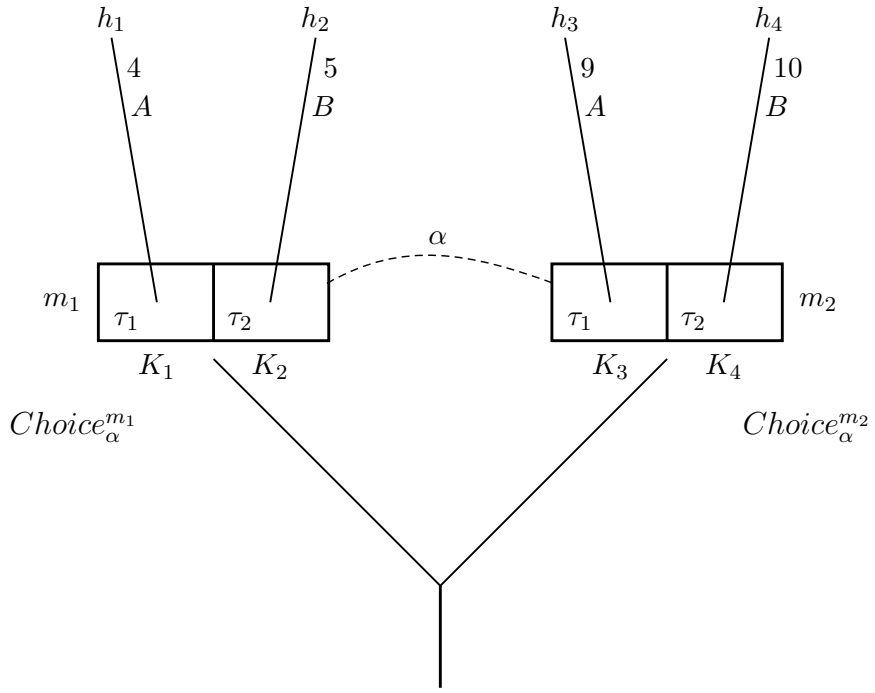

Figure 8: Sure-thing reasoning

each history from $[\tau_1]^I_\alpha = \{h_1, h_3\}$ is at least as valuable as each history from $[\tau_2]^I_\alpha = \{h_2, h_4\}$, since h_1 is less valuable than both h_2 and h_4. The proposed ordering, then, does not support the conclusion that either of the action types τ_1 or τ_2 is better than, or even at least as good as, the other, but instead, that τ_1 and τ_2 are incomparable.

There is, however, a persuasive argument supporting the conclusion that τ_2 should be ordered as preferable to τ_1 on the basis of the information set $I = \{m_1, m_2\}$. After all, although this information does not allow us to determine whether the agent α occupies the moment m_1 or the moment m_2, it does tell us that she occupies one or the other of these moments. So suppose, first, that α occupies m_1. In that case, it is better for her to execute the type τ_2 than the type τ_1, since the token $[\tau_2]^{m_1}_\alpha$ resulting from the execution of τ_2 at the moment m_1 dominates—in the sense of token dominance set out earlier—the token $[\tau_1]^{m_1}_\alpha$, resulting from the execution of τ_1. Next, suppose α occupies m_2. Then again, it is better for her to execute the type τ_2 than the type τ_1, since the token $[\tau_2]^{m_2}_\alpha$ resulting from the execution of τ_2 at the moment m_2 dominates the token $[\tau_1]^{m_2}_\alpha$ resulting from the execution

of τ_1. In each of these two cases, then, it is better for the agent to execute the type τ_2 than the type τ_1, and since these cases exhaust the possibilities provided by the information set I, a pattern of reasoning sometimes described as *sure-thing reasoning* yields the conclusion that τ_2 should be ordered as preferable to τ_1.[22]

What this argument suggests is that, in ranking two action types available to an agent on the basis of an information set, rather than comparing the entire set of outcomes that might, consistent with this information, issue from the execution of each of these action types, we should instead engage in a point-by-point comparison of the action tokens that would result from the execution of the action types at each moment from the information set. In accord with this suggestion, we can now define a dominance ordering on action types in terms of our previous dominance ordering on action tokens, relative to an information set, simply by quantifying over the moments from that information set.

Definition 10 (Dominance orderings on action types; $\preceq^I_\alpha, \prec^I_\alpha$) Let α be an agent from a labeled deontic stit frame, I an information set bearing on α, and τ and τ' action types belonging to $Type^m_\alpha$ for each moment m from I. Then $\tau \preceq^I_\alpha \tau'$ (τ' weakly dominates τ, on the basis of I) if and only if $[\tau]^m_\alpha \leq [\tau']^m_\alpha$ for each moment m from I; and $\tau \prec^I_\alpha \tau'$ (τ' strongly dominates τ, on the basis of I) if and only if $\tau \preceq^I_\alpha \tau'$ and it is not the case that $\tau' \preceq^I_\alpha \tau$.

The reader will note that this procedure for lifting an ordering on action tokens to an ordering on action types mirrors our earlier procedure, codified in Definition 5, for lifting an ordering on histories to an ordering on action tokens. And the current dominance ordering on types likewise inherits the properties of our earlier ordering on tokens: both the strong and weak

[22] This pattern of reasoning is first explicitly characterized as the "sure-thing principle" in Savage [32], but the principle appears already in some of Savage's earlier work, such as [31, p. 58], where he writes concerning situations of uncertainty that "there is one unquestionably appropriate criterion for preferring some act to some others: If for every possible state, the expected income of one act is never less and is in some cases greater than the corresponding income of another, then the former act is preferable to the latter."

orderings on types are transitive, and the strong ordering is irreflexive.

Our dominance ordering on action types can be illustrated by returning to the example from Figure 8. Here, we can see that the action type τ_2 weakly dominates the action type τ_1 on the basis of the information set $I = \{m_1, m_2\}$, since the token that results from executing τ_2 weakly dominates the token that results from executing τ_1 at each moment from this information set: we have $\tau_1 \preceq^I_\alpha \tau_2$, since $[\tau_1]^{m_1}_\alpha \leq [\tau_2]^{m_1}_\alpha$ and $[\tau_1]^{m_2}_\alpha \leq [\tau_2]^{m_2}_\alpha$. On the other hand, τ_1 does not weakly dominate τ_2 since there is some moment from I at which the token that results from executing τ_1 does not weakly dominate the token that results from executing τ_2—indeed, the required token dominance fails at both moments: we do not have $\tau_2 \preceq^I_\alpha \tau_2$, since neither $[\tau_2]^{m_1}_\alpha \leq [\tau_1]^{m_1}_\alpha$ nor $[\tau_2]^{m_2}_\alpha \leq [\tau_1]^{m_2}_\alpha$. Because τ_2 weakly dominates τ_1 on the basis of I but the reverse does not hold, it follows that τ_2 strongly dominates τ_1: we have $\tau_1 \prec^I_\alpha \tau_2$.

Let us now, at last, introduce a new *epistemic ought* operator—written, $\bigodot[\ldots\ kstit\text{: }\ldots]$—allowing for statements of the form

$$\bigodot[\alpha\ kstit\text{: } A].$$

Just like our earlier agentive oughts, an epistemic ought statement of this form can also be taken to mean that the agent α ought to see to it that A, but now in an epistemic sense.

Following the route mapped out in our treatment of agentive oughts, the semantics of epistemic oughts also relies on the idea of optimality—but now of optimal action types, rather than tokens, where the action types that are optimal on the basis of an information set are those that are not strongly dominated, on the basis of that information set.

Definition 11 (Optimal action types; $T\text{-}Optimal^I_\alpha$) Where α is an agent from a labeled deontic stit frame, I is an information set bearing on α, and m is a moment from I,

$$T\text{-}Optimal^I_\alpha = \{\tau \in Type^m_\alpha\ :\ \text{there is no } \tau' \in Type^m_\alpha \text{ such that } \tau \prec^I_\alpha \tau'\}.$$

As with tokens, since the set of types is finite, and because the strong dominance relation is transitive and irreflexive, the set of optimal action types available to an agent must be nonempty.

While an agentive ought of the form $\bigodot[\alpha\ \mathit{stit}\colon A]$ was defined earlier as holding at a moment whenever A is guaranteed by the performance of each action token that is optimal at that moment, a new epistemic ought of the form $\bigodot[\alpha\ \mathit{kstit}\colon A]$ can now be defined as holding at a moment from an information set whenever A is guaranteed, at each moment within that information set, by the execution of each action type that is optimal on the basis of that information set. But which information set? In evaluating an epistemic ought of this form, at a moment m, we focus on the particular information set I_α^m, representing the information available to the agent α herself at the very moment of evaluation. This leads to the following evaluation rule:

Definition 12 (Evaluation rule: $\bigodot[\alpha\ \mathit{kstit}\colon A]$) Where α is an agent and m/h an index from a labeled deontic stit model $\mathcal{M}$,

- $\mathcal{M}, m/h \models \bigodot[\alpha\ \mathit{kstit}\colon A]$ if and only if $[\tau]_\alpha^{m'} \subseteq |A|_{\mathcal{M}}^{m'}$ for each $\tau \in \mathit{T\text{-}Optimal}_\alpha^{I_\alpha^m}$ and for each $m' \in I_\alpha^m$.

It is easy to verify that this epistemic ought, like our earlier agentive ought, is a normal modal operator, that epistemic ought statements are moment determinant, and that it satisfies the very strong deontic principle that, if an agent ought to see to it that A, then that agent has the ability to see to it that A even in the epistemic sense: the formula $\bigodot[\alpha\ \mathit{kstit}\colon A] \supset \Diamond[\alpha\ \mathit{kstit}\colon A]$ is valid.

For illustration, we can return to the abstract example from Figure 8, noting now that, as the diagram indicates, the agent α cannot distinguish between m_1 and m_2, that the statement A holds at m_1/h_1 and m_2/h_3, and that B holds at m_1/h_2 and m_2/h_4. Suppose α occupies the moment m_1 with information set is $I_\alpha^{m_1} = \{m_1, m_2\}$. It is easy to see that τ_2 strongly dominates τ_1 on the basis of this information set—that is, $\tau_1 \prec_\alpha^{I_\alpha^{m_1}} \tau_2$—so that

the agent's sole optimal action type is the unique member of $\textit{T-Optimal}_{\alpha}^{I_{\alpha}^{m_1}} = \{\tau_2\}$. The statement $\bigodot[\alpha\ \textit{kstit}\colon B]$ is therefore settled true at m_1, since the result of executing this optimal action type guarantees the truth of B at each of the two moments from the agent's information set: $[\tau_2]_{\alpha}^{m_1} \subseteq |B|^{m_1}$ and $[\tau_2]_{\alpha}^{m_2} \subseteq |B|^{m_2}$

6.2 Exploring the epistemic ought

The epistemic ought just introduced is proposed as an operator that combines agentive, deontic, and epistemic ideas in the right way, through a formula of the form $\bigodot[\alpha\ \textit{kstit}\colon A]$, to yield an agentive ought statement whose violation invites criticism of the agent. Earlier, we considered the proposal that this idea could be captured simply by appending a knowledge operator to an ordinary agentive ought, through a formula of the form $\mathsf{K}_{\alpha}\bigodot[\alpha\ \textit{stit}\colon A]$. This initial proposal was rejected on the grounds that it offered incorrect predictions in three representative cases. We now return to these three cases to confirm that the current suggestion yields better results.

The first problem was based on the example from Figure 4, later reproduced with action types rendered explicit in Figure 6. As we saw, the problem presented by this example was that, supposing you occupy m_2, the proposal predicts that you know you ought to gamble: $\mathsf{K}_{\alpha}\bigodot[\alpha\ \textit{stit}\colon G]$ is settled true. This statement holds just in case, at each moment indistinguishable for you from m_2, each optimal action token guarantees that you gamble. The indistinguishable moments are m_2 itself and m_3, the optimal action tokens at these moments are $\textit{K-Optimal}_{\alpha}^{m_2} = \{K_3\}$ and $\textit{K-Optimal}_{\alpha}^{m_3} = \{K_7\}$, and both of these action tokens guarantee that you gamble: $K_3 \subseteq |G|^{m_2}$ and $K_7 \subseteq |G|^{m_3}$. Yet this does not seem like the right result—it does not seem like you know you ought to gamble, or that you could reasonably be criticized for failing to gamble.

The current suggestion avoids this first problem, since it does not predict that you ought to gamble, at least in the epistemic sense: $\bigodot[\alpha\ \textit{kstit}\colon G]$ is settled false at m_2. This statement holds just in case, at each moment from your information set $I_{\alpha}^{m_2} = \{m_2, m_3\}$, the

execution of each action type that is optimal on the basis of this information guarantees that you gamble. The set of action types that are optimal on the basis of this information is the entire set $\textit{T-Optimal}_{\alpha}^{I_{\alpha}^{m_2}} = \{\tau_1, \tau_2, \tau_3\}$; each of these action types is optimal because none is even weakly dominated by another—for each, there is some moment within your information set at which the execution of another action type yields a dominating action token. But it is not the case that the execution of each of these optimal action types guarantees that you gamble at each moment from your information set. In particular, the execution of τ_3, the action type of refraining from gambling, at either m_2 or m_3 does not guarantee that you gamble: we have neither $[\tau_3]_{\alpha}^{m_2} \subseteq |G|^{m_2}$ nor $[\tau_3]_{\alpha}^{m_3} \subseteq |G|^{m_3}$.

The second problem for the initial proposal was based on the example from Figure 5, exactly like that depicted in Figures 4 and 6 except that the histories carry different values: this time, since the gamble is peculiar, the histories that result from betting correctly carry no more value than the histories that result from refraining from the gamble. Although action types are not represented explicitly in Figure 5, we can assume that τ_1, τ_2, and τ_3 again represent the action types of betting heads, betting tails, and refraining, and also that, as in Figure 6, the execution of these types result in the respective action tokens K_3, K_4, or K_5 at m_2, and K_6, K_7, or K_8 at m_3.

The problem presented by this example for the initial proposal was that, again supposing that you occupy m_2, the proposal fails to predict that you know you ought not to gamble: $\mathsf{K}_{\alpha}\bigodot[\alpha\ \textit{stit}\colon \neg G]$ is settled false. This statement fails because, at the two moments m_2 and m_3 that you cannot distinguish from m_2, the optimal action tokens are $\textit{K-Optimal}_{\alpha}^{m_2} = \{K_3, K_5\}$ and $\textit{K-Optimal}_{\alpha}^{m_3} = \{K_7, K_8\}$, and not all of these optimal action tokens guarantee that you refrain from gambling: in particular, we have neither $K_3 \subseteq |\neg G|^{m_2}$ nor $K_7 \subseteq |\neg G|^{m_3}$. But contrary to this prediction, it does seem that, in light of your information, that you ought not to gamble, that you know this, or at least that an ideal reasoner would know it, and that you could reasonably be criticized for gambling.

The current suggestion avoids this second problem by correctly predicting that you ought

not to gamble: the statement $\bigodot[\alpha\ kstit\colon \neg G]$ is settled true. On the basis of your information $I_\alpha^{m_2} = \{m_2, m_3\}$, the unique member of the set $\mathit{T\text{-}Optimal}_\alpha^{I_\alpha^{m_2}} = \{\tau_3\}$ is your only available optimal action type, since it strongly dominates each of the others—the execution of τ_3 always yields an action token that weakly dominates the execution of τ_1 or τ_2, and for each, there is some moment in your information set at which the execution of τ_3 strongly dominates. And the execution of this optimal action type guarantees that you refrain from gambling at each moment from your information set: we have both $[\tau_3]_\alpha^{m_2} \subseteq |\neg G|^{m_2}$ and $[\tau_3]_\alpha^{m_3} \subseteq |\neg G|^{m_3}$.

The third problem for the initial proposal was based, once again, on the example from Figures 4 and 6, with the statement letter W, equivalent to the formula $(BH \wedge H) \vee (BT \wedge T)$, now representing the proposition that you win, or bet correctly. In this case, the problem posed for the initial proposal is that, supposing that you occupy m_2, the proposal predicts that you know you ought to win: $\mathsf{K}_\alpha \bigodot[\alpha\ stit\colon W]$ is settled true at m_2. This statement holds because, as we saw, the optimal action tokens available to you at the moments indistinguishable from m_2 are $\mathit{K\text{-}Optimal}_\alpha^{m_2} = \{K_3\}$ and $\mathit{K\text{-}Optimal}_\alpha^{m_3} = \{K_7\}$, and both of these action tokens guarantee that you win: $K_3 \subseteq |W|^{m_2}$ and $K_7 \subseteq |W|^{m_3}$. As we noted earlier, it is not entirely implausible to suppose that you know you ought to win—but not in a sense in which failure to win would invite criticism. You could not legitimately be criticized for failing to win, because winning is not something you are, in the epistemic sense, able to do.

The current suggestion avoids this third problem as well, since it does not predict that you ought to win, at least in the epistemic sense in which failure to win would invite criticism: the statement $\bigodot[\alpha\ kstit\colon W]$ is settled false at m_2. This statement does not hold since the set of optimal action types available to you on the basis of your information set $I_\alpha^{m_2} = \{m_2, m_3\}$ is again the entire set $\mathit{T\text{-}Optimal}_\alpha^{I_\alpha^{m_2}} = \{\tau_1, \tau_2, \tau_3\}$, and it is not the case that the execution of each of these optimal action types guarantees winning at each moment from your information set. In particular, you do not win by executing either τ_2 or τ_3 at m_2, or by executing either τ_1 or τ_3 at m_3: all of $[\tau_2]_\alpha^{m_2} \subseteq |W|^{m_2}$ and $[\tau_3]_\alpha^{m_2} \subseteq |W|^{m_2}$ and $[\tau_1]_\alpha^{m_3} \subseteq |W|^{m_3}$ and $[\tau_3]_\alpha^{m_3} \subseteq |W|^{m_3}$ fail.

Having verified that the epistemic ought operator resolves the problems posed for the initial proposal, it is worth exploring the relations between this new operator and the agentive ought defined earlier. Even though, as we have noted, the epistemic *kstit* operator is strictly stronger than the familiar causal *stit*, it turns out that the epistemic ought is neither stronger nor weaker than the familiar agentive ought: neither of the formulas

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

is valid. A countermodel to the first formula is provided by the example from Figure 5, where, as we have seen, $\textit{T-Optimal}_{\alpha}^{I_{\alpha}^{m_2}} = \{\tau_3\}$, so that $\bigodot[\alpha \; \textit{kstit}\colon \neg G]$ is settled true at m_2, since both $[\tau_3]_{\alpha}^{m_2} \subseteq |\neg G|^{m_2}$ and $[\tau_3]_{\alpha}^{m_3} \subseteq |\neg G|^{m_3}$, but $\textit{K-Optimal}_{\alpha}^{m_2} = \{K_3, K_5\}$, so that $\bigodot[\alpha \; \textit{stit}\colon \neg G]$ is settled false, since it is not the case that $K_3 \subseteq |\neg G|^{m_2}$. What this example shows is that action tokens can be optimal even if they do not result from the execution of optimal action types: here, K_3 is an optimal action token even though it results from the execution of the non-optimal action type τ_1. A countermodel to the second formula is provided by the example from Figures 4 and 6, where $\textit{K-Optimal}_{\alpha}^{m_2} = \{K_3\}$, so that $\bigodot[\alpha \; \textit{stit}\colon BH]$ is settled true at m_2, since $K_3 \subseteq |BH|^{m_2}$, but $\textit{T-Optimal}_{\alpha}^{I_{\alpha}^{m_2}} = \{\tau_1, \tau_2, \tau_3\}$, so that $\bigodot[\alpha \; \textit{kstit}\colon BH]$ is settled false, since it is not the case that $[\tau_2]_{\alpha}^{m_2} \subseteq |BH|^{m_2}$, for example. What this example shows is that an action type can be optimal on the basis of an agent's information even if its execution at some moment consistent with that information does not result in an optimal action token: here, τ_2 is an optimal action type even though the action token K_4 resulting from its execution at m_2 is not optimal.

Although there are not, then, any general connections between the epistemic ought operator defined here and the ordinary agentive ought, the two operators are equivalent in models satisfying the *perfect information* constraint, mentioned earlier, which tells us that an agent always knows which moment she occupies. In this case, the information set for an agent α occupying the moment m is simply $I_{\alpha}^{m} = \{m\}$, from which we can conclude that

$$\textit{K-Optimal}_{\alpha}^{m} = \{[\tau]_{\alpha}^{m} : \tau \in \textit{T-Optimal}_{\alpha}^{I_{\alpha}^{m}}\},$$

or that the optimal action tokens are exactly those resulting from the execution of optimal action types.[23] Given this identity, it follows at once that the two oughts coincide:

$$\bigodot[\alpha \; kstit\text{: } A] \equiv \bigodot[\alpha \; stit\text{: } A].$$

The new epistemic deontic operator introduced here is a conservative extension of the previous causal operator, agreeing if the agent knows everything about the past, leading up to the present moment, disagreeing only in situations in which the agent has some uncertainty about the past, and so about her present situation.

There is one further logical point worth noting: the formula

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

is valid, so that it follows, from the fact that an agent ought to do something in the epistemic sense, that she knows she ought to do it. The current suggestion can therefore be seen as respecting the intuition underlying our initial proposal—that we can be criticized for failing to do what we ought to do only if we know we ought to do it.

7 Generalizations

7.1 Relativism

This section briefly mentions two directions in which our account of epistemic oughts can be elaborated, beginning with a generalization in the direction of relativism, or assessment sensitivity.

Stepping back a bit: We have now considered two kinds of agentive ought statements. The first is the ordinary agentive ought, carried by formulas of the form $\bigodot[\alpha \; stit : \; A]$, governed by the evaluation rule from Definition 7; the second is the epistemic ought, carried by formulas of the form $\bigodot[\alpha \; kstit\text{: } A]$, governed by the evaluation rule from Definition 12.

[23] Or equivalently, looked at from the other side, we have $T\text{-}Optimal_\alpha^{I^m} = \{Label(K) : K \in K\text{-}Optimal_\alpha^m\}$, so that the optimal action types are exactly those that are labels of optimal action tokens.

Although I have not used this language, it is natural to speak of the contrast between these two kinds of ought statements as the contrast between *objective* and *subjective* oughts—between statements describing what the agent ought to do on the basis of the actual facts, regardless of her information about these facts, and statements describing what the agent ought to do only on the basis of her own information. In the example from Figures 4 and 6, for instance, we could speak of $\bigodot[\alpha \; \mathit{stit}\colon BH]$ as an objective ought statement, settled true at m_2 because betting heads is what you actually ought to do if you occupy that moment, regardless of your information about the moment you occupy, and we could speak of $\bigodot[\alpha \; \mathit{kstit}\colon BH]$ as a subjective ought statement, settled false at m_2 because your limited information about the moment you occupy does not support the conclusion that you ought to bet heads.

Even though this way of speaking may be natural, however, it faces a significant objection. The objection is that, in advancing the idea that there are distinct objective and subjective agentive ought statements—formed from different operators, governed by different evaluation rules—we are, in effect, supposing that we have discovered two different senses, or meanings, in the word "ought." Of course, philosophy often proceeds like this, by discovering hidden ambiguities in items of ordinary language, which are then teased apart and provided with different formal explications. But in this case, it may seem to be forced, or artificial, to imagine that our different ways of understanding agentive ought statements depend on a lexical ambiguity in the word "ought."

This objection has been developed with great force by John MacFarlane in Chapter 11 of his recent [26], where he argues that, if "ought" carries separate objective and subjective meanings, then it is hard to understand certain kinds of moral or prudential disagreements. Returning to our central example, suppose that I, knowing full well that I placed the coin heads up, and so knowing that you occupy the moment m_2, state that you ought to bet heads, but that you, without knowing whether you occupy m_2 or m_3, deny that you ought to bet heads. Then, apparently, my statement would carry the objective sense of "ought," since

it is based on the actual facts about your situation, while yours would carry the subjective sense, since it is based only on your less specific information about the situation you occupy. If our statements carry these different senses of the word "ought"—if we are using the word in these different ways—then it is hard to understand how we could be disagreeing at all, as opposed to simply talking past each other. But in fact, we do seem to be disagreeing—we seem to be disagreeing about what you ought to do.

In addition to developing this objection to the idea that ought statements carry distinct meanings, objective and subjective, MacFarlane also offers an explanation of the apparent contrast between objective and subjective oughts. His proposal is that agentive ought statements are interpreted relative to a body of information determined by the context from which they are assessed, and that what might appear to be different objective and subjective meanings of these statements result, instead, from different relations between the information available at assessment and the information available to the subjects of these oughts. More specifically, MacFarlane suggests that ought statements have a more "objective feel" when the information on the basis of which they are assessed is more accurate than the information available to the subject of the ought, and a more "subjective feel" when the two bodies of information are more closely matched.[24]

MacFarlane presents this suggestion in the course of developing a sophisticated relativist semantic picture, which I cannot describe in any detail here. I do want to show, however, how certain aspects of his suggestion can be modeled within the current framework. In particular, I want to show how the distinction between objective and subjective agentive oughts can be reconstructed, without postulating lexical ambiguity, by extending our current indices with an additional parameter, representing the information on the basis of which agentive ought statements are evaluated.

We begin, then, by defining an *informationally extended index* as a triple of the form $m/h/I$, with m/h an ordinary index and I an information set containing the moment m.[25]

[24] See Section 11.3 of MacFarlane [26].

[25] We focus here only on rules for evaluating formulas at informationally extended indices like these, without

Where $\mathcal{M}$ is a labeled deontic stit model that results from supplementing a labeled deontic stit frame with the valuation v, we can then define its corresponding *informationally extended labeled deontic stit model* $\mathcal{M}'$ as the model that results from supplementing the same labeled deontic stit frame with the new valuation v', where, for each propositional constant p from the background language, $v(p)$ now consists of all indices of the form $m/h/I$ that extend the ordinary indices of the form m/h from $v(p)$.

Since the additional parameter from an extended index is supposed to represent the information on the basis of which statements are evaluated, what it means to say that the extended index $m/h/I$ from the model $\mathcal{M}'$ satisfies a statement A—formally, that $\mathcal{M}', m/h/I \models A$—is that A holds at the ordinary index m/h on the basis of the information from I. In the same way, we can say that A is *settled true* on the basis of I at the moment m from $\mathcal{M}'$ just in case $\mathcal{M}', m/h/I \models A$ for each h from H^m, *settled false* on the basis of I just in case $\mathcal{M}', m/h/I \models \neg A$ for each h from H^m, and *moment determinant* on the basis of I just in case it is either settled true or settled false. And we can define the proposition expressed by a sentence A at a moment m on the basis of the information from I from $\mathcal{M}'$ as the set $|A|^{m,I}_{\mathcal{M}'} = \{h \in H^m : \mathcal{M}', m/h/I \models A\}$, containing the histories through m along which the sentence A is satisfied on the basis of the information from I.

These ideas can all be defined formally for statements from the existing language, in recursive fashion, simply by adapting each of the evaluation rules set out thus far so that: (i) reference to the ordinary model $\mathcal{M}$ with valuation v is replaced with reference to the corresponding informationally extended model $\mathcal{M}'$ with valuation v', (ii) reference to the ordinary index m/h from $\mathcal{M}$ is replaced with reference to the extended index $m/h/I$ from $\mathcal{M}'$, and (iii) reference to the ordinary proposition $|A|^m_{\mathcal{M}}$ is replaced with reference to the information relative proposition $|A|^{m,I}_{\mathcal{M}'}$. The results of this exercise are unsurprising. Since

worrying about the ways in which these extended indices should be related to contexts of language use or assessment. We thus focus on what MacFarlane refers to as *semantics*, rather than *post-semantics*; see Section 3.2 of MacFarlane [26]. Readers seeking further illumination can consult Belnap [5] for a discussion of *pre-semantics*.

the existing evaluation rules do not draw on the informational component of an extended index in any way, simply adjusting these rules to apply in the presence of this additional component does not make any real difference. This point can be put precisely by noting, where $\mathcal{M}'$ is an extended model corresponding to the ordinary model $\mathcal{M}$, and A is a statement from our existing language, that

$$\mathcal{M}', m/h/I \models A \text{ just in case } \mathcal{M}, m/h \models A;$$

that is, A holds at an index from an ordinary model just in case it holds at any informational extension of that index from the corresponding extended model.

What the extended setting does allow, however, is the introduction of operators that draw on the informational component of an extended index. And in particular, we can now introduce the new *informational ought* operator—written, $\bigodot[\ldots\ istit\text{: }\ldots]$—allowing statements of the form

$$\bigodot[\alpha\ istit\text{: } A]$$

again meaning that α ought to see to it that A, governed by the evaluation rule:

Definition 13 (Evaluation rule: $\bigodot[\alpha\ istit\text{: } A]$) Where α is an agent, I is an information set bearing on α, and $m/h/I$ is an index from an extended labeled deontic stit model $\mathcal{M}'$,

- $\mathcal{M}', m/h/I \models \bigodot[\alpha\ istit\text{: } A]$ if and only if $[\tau]^{m'}_{\alpha} \subseteq |A|^{m',I}_{\mathcal{M}}$ for each $\tau \in \textit{T-Optimal}^{\,I}_{\alpha}$ and for each $m' \in I$.

As the reader can see, this evaluation rule is exactly like the rule from Definition 12, governing the epistemic ought, except that, rather than drawing on the information set I^m_α, representing the information available to the subject of the ought, it draws on the information set I from the extended index, on the basis of which the ought is evaluated.

Like our other ought operators, this informational ought is a normal modal operator, supporting ought statements that are moment determinant. We are not in a position, however, to arrive at a sensible formulation of the agentive deontic principle that ought implies

ability. It is easy to see, for example, that the statement $\bigodot[\alpha\ \mathit{istit}\colon A] \supset \Diamond[\alpha\ \mathit{kstit}\colon A]$ is invalid, since its consequent draws on the agent's own information while its antecedent draws on the information available at the point of assessment, which might well be more accurate.[26]

What is the relation between an informational ought statement $\bigodot[\alpha\ \mathit{istit}\colon A]$ and our two previous ought statements, the ordinary agentive ought $\bigodot[\alpha\ \mathit{stit}\colon A]$ and the epistemic ought $\bigodot[\alpha\ \mathit{kstit}\colon A]$? The answer is that the informational ought behaves like an ordinary agentive ought when it is evaluated at an index of the form $m/h/I^*$, where $I^* = \{m\}$ is the information set containing perfect information about the moment occupied by the agent, and that it behaves like an epistemic ought when it is evaluated at an index of the form $m/h/I^m_\alpha$, where I^m_α is the information set matching the agent's own information. Again, the point can be put precisely by noting, where $\mathcal{M}'$ is an extended labeled deontic stit model, that

$$\mathcal{M}', m/h/I^* \models \bigodot[\alpha\ \mathit{istit}\colon A] \text{ just in case } \mathcal{M}', m/h/I^* \models \bigodot[\alpha\ \mathit{stit}\colon A],$$
$$\mathcal{M}', m/h/I^m_\alpha \models \bigodot[\alpha\ \mathit{istit}\colon A] \text{ just in case } \mathcal{M}', m/h/I^m_\alpha \models \bigodot[\alpha\ \mathit{kstit}\colon A].$$

The first of these equivalences follows from the correspondence, noted in the previous section, between optimal action tokens and instances of action types that are optimal under perfect information; the second follows immediately, since the evaluation rules for the epistemic and informational ought operators coincide when the informational ought is evaluated on the basis of the agent's own information.

Combining these equivalences with the previously displayed fact, we can conclude, where $\mathcal{M}'$ extends the ordinary labeled deontic stit model $\mathcal{M}$, and A is a statement from the original language, that

[26] It seems that an appropriate deontic principle for the informational ought would require the introduction of an informational possibility operator $\Diamond_i$ together with an information version of the *kstit* operator, allowing us to evaluate statements of the form $[\alpha\ \mathit{istit}\colon A]$ in isolation. With operators like these in place, the deontic principle could then be formulated as $\bigodot[\alpha\ \mathit{istit}\colon A] \supset \Diamond_i[\alpha\ \mathit{istit}\colon A]$, but the definition of these new informational operators will not be considered here.

$\mathcal{M}', m/h/I^* \models \bigodot[\alpha\ istit{:}\ A]$ just in case $\mathcal{M}, m/h \models \bigodot[\alpha\ stit{:}\ A]$,

$\mathcal{M}', m/h/I^m_\alpha \models \bigodot[\alpha\ istit{:}\ A]$ just in case $\mathcal{M}, m/h \models \bigodot[\alpha\ kstit{:}\ A]$.

And then, if we again consider the ordinary ought as objective and the epistemic ought as subjective, we can see a way of reconstructing something like MacFarlane's suggestion in the present setting: if we work from a standpoint that takes the informational ought as fundamental, then what appears to be an objective agentive ought statement can be understood as an informational ought based on perfect information about the situation confronting the agent of the ought, and what appears to be a subjective ought statement can be understood as an informational ought based only on the information available to the agent herself.

As MacFarlane goes on to emphasize, agentive ought statements can be evaluated on the basis of information that is neither perfect nor equivalent to the agent's own information, leading to oughts that are neither objective nor subjective—there are as many ways of understanding agentive ought statements as there are information sets on the basis of which they can be evaluated.

This point, too, can be illustrated with the current informational ought operator, as we can see by considering the new situation depicted in Figure 9. Here, we are to imagine that a pea has been placed beneath one of three shells—left, center, or right—and that you then bet on its location, winning a dollar if you are correct. Your choice occurs at m_1 if the coin was placed under the left shell, at m_2 if the coin was placed under the center shell, and at m_3 if the coin was placed under the right shell, though you do not know which shell hides the coin, and so you do not know which moment you occupy. The action types available to you are τ_1, τ_2, and τ_3, representing the actions of betting that the pea is under the left, center, or right shells, respectively. The execution of τ_1 thus leads to the performance of K_1, K_4, or K_7, depending on whether you occupy m_1, m_2, or m_3; likewise, the execution of τ_2 leads to the performance of K_2, K_5, or K_8, and the execution of τ_3 leads to the performance of K_3, K_6, or K_9. The histories h_1, h_5, and h_9, in which you bet correctly, receive a value of 1,

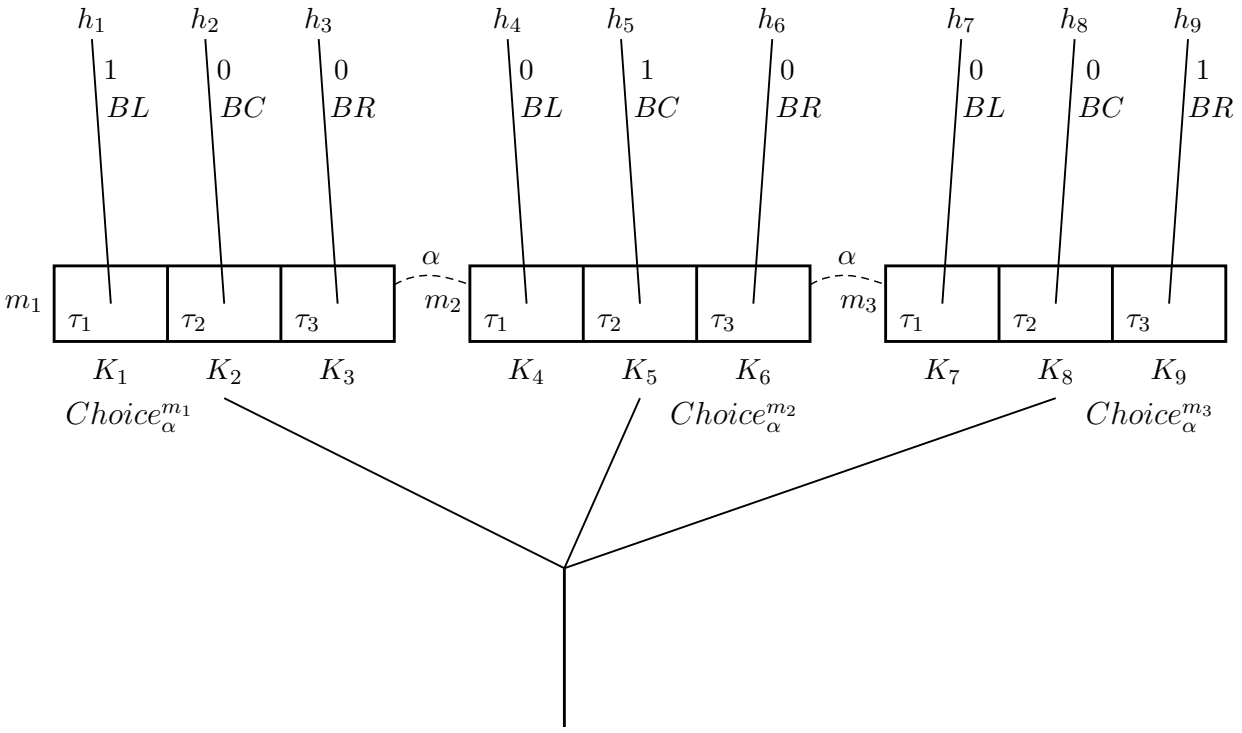

Figure 9: A shell game

while the rest have a value of 0. The statements BL, BC, and BR stand for the statements that you guess that the coin is under the left, center, or right shell, respectively; the first holds at m_1/h_1, m_2/h_4, and m_3/h_7, while the second holds at m_1/h_2, m_2/h_5, and m_3/h_8, and the third holds at m_1/h_3, m_2/h_6, and m_3/h_9. Finally, we take α to represent you, $\sim_\alpha$ to represent indistinguishability from your perspective, and we suppose that, in fact, you occupy the moment m_1—the pea lies under the left shell.

Let us now consider what you ought to do at the moment you occupy from the standpoint of three different information sets. The first is the set $I^* = \{m_1\}$ representing perfect information about the moment you occupy. The second is the set $I_\alpha^{m_1} = \{m_1, m_2, m_3\}$ representing your own information, and reflecting your own complete uncertainty about the moment you occupy. The third is the set $I' = \{m_1, m_2\}$ representing the information that you occupy either m_1 or m_2, without specifying which—this information is better than your own, but still not perfect.

Beginning with optimality, it is easy to verify that:

$$\begin{aligned} \textit{T-Optimal}_{\alpha}^{I^*} &= \{\tau_1\}, \\ \textit{T-Optimal}_{\alpha}^{I'} &= \{\tau_1, \tau_2\}, \\ \textit{T-Optimal}_{\alpha}^{I_\alpha^{m_1}} &= \{\tau_1, \tau_2, \tau_3\}. \end{aligned}$$

In other words: on the basis of the information set I^*, according to which the pea lies under the left shell, the action type of betting left is optimal; on the basis of the information set I', according to which the pea lies under either the left shell or the center shell, without further specification, the action types of betting left and betting center are both optimal; and on the basis of the information $I_\alpha^{m_1}$, which leaves the location of the pea entirely open, each of the available action types is optimal.

As we have seen, informational oughts evaluated on the basis of the information set I^*, representing perfect information, have the character of objective ought statements, while the informational oughts evaluated on the basis of the information set $I_\alpha^{m_1}$, representing your own information, have the character of subjective oughts. But what about informational oughts evaluated on the basis of the I', an information set that is neither perfect nor a representation of your own information? We note, first of all, that the statement

$$\bigodot[\alpha \; \textit{istit}\text{: } BL \vee BC] \wedge \neg \bigodot[\alpha \; \textit{istit}\text{: } BL]$$

is settled true at m_1 on the basis of I'. The first conjunct holds because both of the action types that are optimal on the basis of I'—both τ_1 and τ_2—support betting left or betting center at each moment from the information set: we have both $[\tau_1]_\alpha^m \subseteq |BL \vee BC|^{m,I'}$ and $[\tau_2]_\alpha^m \subseteq |BL \vee BC|^{m,I'}$ where m is m_1 or m_2. The second conjunct holds because it is not the case that each of these optimal action types supports betting left: in particular, we do not have $[\tau_2]_\alpha^m \subseteq |BL \vee BC|^{m,I'}$ for m either m_1 or m_2.

But the displayed statement does not hold when it is evaluated on the basis of I^*. In this case, the second conjunct fails, since the unique action type that is optimal on the basis of I^*—that is, τ_1—does support betting left at the unique moment from that information set:

we have $[\tau_1]^{m_1}_{\alpha} \subseteq |BL|^{m_1,I^*}$. Nor does the displayed statement hold when it is evaluated on the basis of $I^{m_1}_{\alpha}$. In this case, the first conjunct fails, since it is not the case that each of the action types that are optimal on the basis of $I^{m_1}_{\alpha}$—that is, each of τ_1, τ_2, and τ_3—supports betting left or betting center at each moment from this information set: in particular, we do not have $[\tau_3]^{m}_{\alpha} \subseteq |BL \vee BC|^{m,I^{m_1}_{\alpha}}$ for m either m_1 or m_2 or m_3.

Since the displayed statement holds when evaluated on the basis of I' but does not hold when evaluated on the basis of I^* or $I^{m_1}_{\alpha}$, supporting oughts of objective or subjective characters, it follows I' supports a pattern of ought statements that is neither objective nor subjective.

7.2 Conditional oughts

Still working in the extended informational setting, we next consider how the present account of ought statements can be generalized to a rudimentary treatment of conditional oughts. Part of what makes the treatment rudimentary is that it allows conditionalization only on moment determinant statements from the original language—statements that are always, at any moment, either settled true or settled false.

As we recall from our initial discussion of the concept, the proposition expressed by a sentence A is, in the most global sense, the entire set $|A| = \{m/h : m/h \models A\}$ of indices at which that sentence is true. If A happens to be moment determinant, then this proposition has the property that, whenever it contains any index of the form m/h, it also contains every index of the form m/h', for each history h' from H^m. In that case, since the indices belonging to the proposition expressed by A are entirely determined by the set of moments from those indices, the proposition itself carries no more information than the set of moment at which A is settled true, and can therefore be represented by that set.

To formulate this idea precisely, and in the informational setting, we now stipulate that, where A is a moment determinant sentence, the *settled proposition* expressed by A on the

basis of the information set I in the extended model $\mathcal{M}'$ is the set

$$|A|^{\Box,I}_{\mathcal{M}'} = \{m : \mathcal{M}', m/h/I \models A \text{ for each } h \text{ from } H^m\}$$

of moments at which that sentence is settled true on the basis of I. Of course, since our attention is restricted to statements from the original language, where the informational component of an index is idle, we can safely omit mention of this component, referring to a settled proposition of this kind as $|A|^{\Box}_{\mathcal{M}'}$, or even more simply, when the background model can be taken for granted, as $|A|^{\Box}$.

This notation can be illustrated by returning to the example from Figures 4 and 6 and considering the statement H, carrying the meaning that I have placed the coin heads up. Here, the global proposition expressed by H is the set $|H| = \{m_2/h_1, m_2/h_2, m_2/h_3\}$, containing all the indices at which this sentence is true; the settled proposition expressed by H is the set $|H|^{\Box} = \{m_2\}$, containing the unique moment at which the sentence is settled true.

We can now introduce a three-place *conditional informational ought* operator—written, $\bigodot([\ldots \mathit{istit}\text{:}\ldots] \,/\, \ldots)$—allowing construction of statements of the form

$$\bigodot([\alpha \mathit{istit}\text{:}\, A] \,/\, B)$$

to express the idea: if B, the agent α ought to see to it that A. In the current setting, we interpret the antecedent of a conditional ought like this as restricting the information on the basis of which the ought statement is evaluated. Our evaluation rule can therefore be arrived at simply by modifying the rule for ordinary informational oughts, set out in Definition 13, so that the information set drawn from the extended index is intersected with the settled proposition expressed by the antecedent of the conditional.

Definition 14 (Evaluation rule: $\bigodot([\alpha \mathit{istit}\text{:}\, A] \,/\, B)$) Where α is an agent, I is an information set bearing on this agent, $m/h/I$ is an index from an extended labeled deontic stit model $\mathcal{M}'$, and B is sentence that is moment determinant in that model,

- $\mathcal{M}', m/h/I \models \bigodot([\alpha \ \mathit{istit}: A] \,/\, B)$ if and only if $[\tau]^{m'}_{\alpha} \subseteq |A|^{m'}_{\mathcal{M}'}$ for each $\tau \in \mathit{T\text{-}Optimal}^{I \cap |B|^{\Box}_{\mathcal{M}'}}_{\alpha}$ and for each $m' \in I \cap |B|^{\Box}_{\mathcal{M}'}$.

Inspection of this rule reveals the technical reason for limiting the antecedents of conditional oughts to moment determinant statements. Since an information set is defined as a set of moments, rather than a set of moment/history pairs, any further information that refines such a set must be of the same type—also a set of moments, or a settled proposition. And it is only moment determinant statements whose meanings can be represented by settled propositions, rather than propositions of the more usual sort, containing full indices.[27]

The conditional informational ought defined here is a normal conditional operator, and conditional ought statements, like all of our deontic statements, are moment determinant. Furthermore, if we take $\top$ as a trivial statement, settled true at every moment, then it is easy to see that the statement $\bigodot([\alpha\ \mathit{istit}\text{: } A] \,/\, \top)$ is equivalent to the statement $\bigodot[\alpha\ \mathit{istit}\text{: } A]$, so that our account of conditional informational oughts generalizes our earlier treatment of ordinary informational oughts.

The new operator can be illustrated by returning once again to the example from Figures 4 and 6, supposing that you occupy m_2, and evaluating oughts on the basis of the information set $I = \{m_2, m_3\}$. Here, our treatment of informational oughts happily fails to predict that you ought to gamble: the statement $\bigodot[\alpha\ \mathit{istit}\text{: } G]$ is settled false at m_2 on the basis of I, because it is not the case that, at each moment from I, the execution of each member of the set $\mathit{T\text{-}Optimal}^I_\alpha = \{\tau_1, \tau_2, \tau_3\}$ of action types that are optimal on the basis of I guarantees that you gamble. But our account of conditional oughts does allow us to conclude that you ought to gamble if I have placed the coin heads up: the statement $\bigodot([\alpha\ \mathit{istit}\text{: } G] \,/\, H)$ is settled true at m_2 on the basis of I. Why? Because the settled proposition expressed by the statement that I have placed the coin heads up is $|H|^\Box = \{m_2\}$, so that the original information set restricted by this settled proposition is $I \cap |H|^\Box = \{m_2\}$. On the basis of this new information set, the unique member of $\mathit{T\text{-}Optimal}^{I \cap |H|^\Box}_\alpha = \{\tau_1\}$ is your only optimal action type, since it strongly dominates the others—at the moment m_2, where the coin has

[27] These restrictions—even the restriction of the agent's information to a set of moments, rather than a full proposition—are all driven by considerations of simplicity; they could be relaxed, leading to a more expressive formalism, but only at the cost of more complexity than would be tolerable in this paper.

been placed heads up, betting heads yields better results than betting tails or refraining. And the execution of this unique optimal action type at m_2 guarantees that you gamble: $[\tau_1]_\alpha^{m_2} \subseteq |G|^{m_2}$.

Although we cannot discuss the logic of this conditional ought operator here in any detail, it is useful to show how it allows us to invalidate three problematic patterns of inference in conditional deontic logic. The first of these is factual detachment: the inference from premises of the form $\bigodot([\alpha\ \mathit{istit}\colon A] \,/\, B)$ and B to a conclusion of the form $\bigodot[\alpha\ \mathit{istit}\colon A]$. Here, a counterexample can be found by reflecting on the scenario just above, which supports the premise that you ought to gamble if I have placed the coin heads up: $\bigodot([\alpha\ \mathit{istit}\colon G] \,/\, H)$ is settled true at m_2 on the basis of I. And of course, I did place the coin heads up: H is settled true at m_2. But again, it does not follow that you ought to gamble: $\bigodot[\alpha\ \mathit{istit}\colon G]$ is settled false.

The second problematic inference is antecedent strengthening, or monotonicity: the inference from a premise of the form $\bigodot([\alpha\ \mathit{istit}\colon A] \,/\, B)$ to a conclusion of the form $\bigodot([\alpha\ \mathit{istit}\colon A] \,/\, B \wedge C)$. This inference can be invalidated even though our analysis of conditional oughts avoids any appeal to the kind of similarity relations among indices of evaluation that underlie so many conditional deontic logics, as well as conditional logics more generally. A counterexample can be found by returning to our earlier Figure 5, depicting the peculiar gamble, and evaluating oughts at m_2 on the basis of $I = \{m_2, m_3\}$. Here, we have the premise that you ought not to gamble: $\bigodot[\alpha\ \mathit{istit}\colon \neg G]$ is settled true at m_2 on the basis of I, since the unique member of your set $\mathit{T\text{-}Optimal}_\alpha^I = \{\tau_3\}$ of optimal actions on the basis of I guarantees that you refrain from gambling at both moments from this set: $[\tau_3]_\alpha^{m_2} \subseteq |\neg G|^{m_2}$ and $[\tau_3]_\alpha^{m_3} \subseteq |\neg G|^{m_3}$. From $\bigodot[\alpha\ \mathit{istit}\colon \neg G]$, we have $\bigodot([\alpha\ \mathit{istit}\colon \neg G] \,/\, \top)$. But we cannot conclude that you ought not to gamble if the coin is placed heads up, since the set $\mathit{T\text{-}Optimal}_\alpha^{I \cap |H|^\Box} = \{\tau_1, \tau_3\}$ of action types that are optimal on the basis of $I \cap H^\Box$ contains τ_1, whose execution guarantees that you gamble. The statement $\bigodot([\alpha\ \mathit{istit}\colon \neg G] \,/\, H)$ is therefore settled false at m_2 on the basis of I, and from this, since H is logically equivalent

to $\top \wedge H$, the normality properties of conditional modal logics allow us to conclude that $\bigodot([\alpha\ \mathit{istit}\colon \neg G] \,/\, \top \wedge H)$ is settled false as well.

The third problematic inference is a form of reasoning by cases: the inference from premises of the form $\bigodot([\alpha\ \mathit{istit}\colon A] \,/\, B)$ and $\bigodot([\alpha\ \mathit{istit}\colon A] \,/\, C)$, together with the disjunction $B \vee C$, to a conclusion of the form $\bigodot[\alpha\ \mathit{istit}\colon A]$. To establish the invalidity of this inference, we return to our central example from Figures 4 and 6, evaluating formulas at m_2 on the basis of $I = \{m_2, m_3\}$. We have already seen, in this example, that you ought to gamble if I have placed the coin heads up: $\bigodot([\alpha\ \mathit{istit}\colon G] \,/\, H)$ is settled true at m_2 on the basis of I. In just the same way, we can conclude that you ought to gamble if I have placed the coin tails up: $\bigodot([\alpha\ \mathit{istit}\colon G] \,/\, T)$ is settled true at m_2 on the basis of I because the execution of the unique member of the set $\mathit{T\text{-}Optimal}_{\alpha}^{I \cap |T|^{\Box}} = \{\tau_2\}$ of actions that are optimal on the basis of $I \cap |T|^{\Box} = \{m_3\}$ guarantees that you gamble at the unique moment from this information set: $[\tau_2]_{\alpha}^{m_3} \subseteq |G|^{m_3}$. And of course, we have the additional premise that I have placed the coin either heads up or tails up: the statement $H \vee T$ is likewise settled true at m_2. Yet, again, we cannot conclude that you ought to gamble: the statement $\bigodot[\alpha\ \mathit{istit}\colon G]$ is settled false.

It is worth pointing out that this central example, from Figures 4 and 6, is closely related to two previous examples from the literature, both deployed to challenge the validity of reasoning by cases. The first is the example depicted in Figure 5.3 of [18]. That earlier example is exactly like the current example from Figures 4 and 6 both in its interpretation of the actions available to the agents and in the values it assigns to outcomes, differing only in temporal sequence: while, in the current example, I place the coin on the table heads up or tails up before you face your choice to bet heads, bet tails, or refrain, in the previous example, our choices are simultaneous—I place the coin on the table at the same moment that you make your bet. Although these two situations are very similar—they would collapse into the same normal-form game, for instance—the temporal difference is crucial in the framework of branching time. In the earlier example, it is still indeterminate at the

moment you place your bet whether the coin is heads up or tails up, while in the current example, the status of the coin is settled, though you are uncertain what that status is. The approach set out here—not just in the current sketch of conditional oughts, but throughout the present paper—involves adapting ideas from the earlier treatment of ought statements in the presence of metaphysical indeterminacy to apply to epistemic uncertainty as well.[28]

The second point of contact with the previous literature involves an example introduced by Donald Regan [30], discussed by Derek Parfit in unpublished work [27], and then adapted to form a puzzle in deontic logic by Niko Kolodny and MacFarlane in a paper [23] that has given rise to an extensive secondary literature.[29] The example is this: A group of ten miners has, at some point in the past, entered either shaft A or shaft B, though we do not know which. Flood waters are rising and we have enough sandbags to block one of the shafts, keeping the water out, but not both. If we do block one of the shafts, we save all ten miners in case they happen to be in the blocked shaft, but then all the water flows into the shaft that is not blocked, rising to a level that it kills all ten miners in case they happen to be in that shaft. If we refrain from blocking either shaft, the water will flow evenly into both, rising to a level that it kills only a single miner in whichever shaft the group of miners is located. The puzzle this example presents for Kolodny and MacFarlane is that they would like to accept both (i) If the miners are in shaft A, we ought to block shaft A, and (ii) If the miners are in shaft B, we ought to block shaft B. And of course the situation supports (iii) The miners are in shaft A or the miners are in shaft B. From (i) through (iii), reasoning by cases, in some form or other, seems to yield (iv) We ought to block shaft A or we ought to block shaft B. But Kolodny and MacFarlane reject (iv) as contrary to our moral intuitions,

[28] In particular, the discussion in Section 6.1 of the current paper leading up to the dominance ordering on action types parallels the discussion from Section 4.1 of [18].

[29] Highlights from this literature, for me personally, include Cariani, Kaufmann, and Kaufmann [8], Carr [9], Charlow [10], Dowell [13], and Willer [39]. Since my limited goal here is only to show that Kolodny and MacFarlane's example fits naturally into the framework of stit semantics, I cannot compare my treatment to the very different approaches taken in these papers.

and instead favor (v) We ought to block neither shaft.

Once again, this miners example is structurally similar to our central example from Figures 4 and 6, as we can see by reinterpreting that diagram. This time, instead of taking α and β to represent you and me, we take α to represent the group of us who must decide what to do for the miners and β to represent the group of miners themselves.[30] The moment m_1 represents the point at which the miners decide to enter either shaft A, performing K_1, or shaft B, performing K_2. After the miners decide whether to enter shaft A or shaft B, as the waters rise, our decision takes place either at m_2 or at m_3, which we cannot distinguish since we do not know which shaft the miners entered. The options before us—blocking shaft A, blocking shaft B, or refraining from blocking either shaft—are represented by the action types τ_1, τ_2, and τ_3, whose execution at m_2 or m_3 gives rise to the action tokens K_3 through K_8, as the diagram indicates. In order to accommodate our new reading of the diagram, the sentence letters must be provided with different interpretations: H and T should now be taken to mean that the miners entered shaft A or shaft B, respectively; BH and BT should be taken to mean that we block shaft A or shaft B, respectively; and G, still equivalent to $BH \vee BT$, should be taken to mean that we block one shaft or the other. Finally, if we let the value of a history reflect the number of miners saved from drowning in that history, the values of h_1, h_2, h_4, and h_5 can remain unchanged, but h_3 and h_6 must now be assigned the value 9, rather than 5.

This adjustment in the value of h_3 and h_6 is the only change to the situation depicted, and note that even this change does not affect the ordinal value relations among the different histories: h_3 and h_6 are still better than h_2 and h_4 but worse than h_1 and h_5. Since our treatment of oughts depends only on these ordinal relations among outcome values, exactly the same statements hold at the same indices after the adjustment as before. In particular,

[30] One of the chief advantages of stit semantics is that it allows for a careful logical analysis of group actions in terms of the actions taken by the individuals that constitute those groups, but in this case there is no need for that level of detail; here, it is easiest simply to treat both the group of miners and the group of potential helpers as individuals.

taking $I = \{m_2, m_3\}$ as the information set on the basis of which statements are evaluated, we now have $\bigodot([\alpha \; istit\text{: } BH] \,/\, H)$ and $\bigodot([\alpha \; istit\text{: } BT] \,/\, T)$ settled true at any moment from that information set. These two statements are plausible representations of (i) and (ii) above: that we ought to block shaft A if the miners are in shaft A, and that we ought to block shaft B if the miners are in shaft B. And of course, we also have $H \vee T$, representing (iii): that the miners are in shaft A or shaft B. But the statement $\bigodot[\alpha \; istit\text{: } BH] \vee \bigodot[\alpha \; istit\text{: } BT]$ is settled false, so that we are not forced to conclude (iv): that we ought to block shaft A or we ought to block shaft B.

The present treatment of conditional oughts, then, allows us to avoid the inference that Kolodny and MacFarlane were concerned to avoid, from (i) through (iii) to (iv). There are, however, a number of important differences between the present treatment and that suggested by Kolodny and MacFarlane. I will mention only three. First, Kolodny and MacFarlane want to represent conditional oughts through the combination of an indicative conditional together with a separate ought operator, rather than through a fused conditional ought operator, as in the present treatment. This is a significant difference, but I am not sure that it is insurmountable. It would be an interesting project to try to reconstruct the present treatment by introducing a new indicative conditional into the current framework, which would then modify the information set on the basis of which an ordinary informational ought is evaluated. Second, Kolodny and MacFarlane's analysis is based on an ordering on outcomes, or worlds, that is defined as relative to an information set, possibly varying from one information set to another.[31] By contrast, the present treatment is based on a value assignment to outcomes, or histories, that is fixed in advance. What varies from one information set to another is not the primitive value assignment to histories, but the derived ordering on action types—while the action types of blocking shaft A and blocking shaft B are incomparable based on no information about the location of the miners, for example, the

[31] They refer to this property of the outcome ordering as *serious information dependence*; see Kolodny and MacFarlane [23, p. 133 ff.].

action type of blocking shaft A dominates the action type of blocking shaft B on the basis of the information that the miners are in shaft A.

Finally, Kolodny and MacFarlane are concerned, not only to avoid the conclusion (iv), but to reach the conclusion (v): that we ought to block neither shaft. The present treatment does not support this conclusion—the statement $\bigodot[\alpha \; istit\text{: } \neg(BH \vee BT)]$ is settled false—and in many ways, that seems like the right result to me. I can understand the intuition behind (v) in the situation exactly as Kolodny and MacFarlane describe it: Why block a shaft, risking the lives of all nine miners, for the chance of saving only one more miner than we could save by blocking neither shaft? But that intuition seems to hinge on cardinality comparisons between the values of different outcomes that I do not know how to account for in any way that is not arbitrary. Consider situations in which outcomes are assigned different cardinal values, while still respecting the original ordinal relations among outcome values. Imagine, for example, that, with neither shaft blocked, the water would rise high enough to drown nine miners—so that, in this new situation, h_3 and h_6 would receive the value 1, rather than 9. Would it still seem wrong to block a shaft, risking the life of a single miner for the chance of saving nine more? What if the unblocked waters would rise high enough to drown eight miners? Is it worth risking the lives of two for the chance of saving eight more? Is it worth risking three for the chance of saving seven more? My intuitions fail.

8 Conclusion

This paper proposes one way in which agentive, deontic, and epistemic concepts might combine to form epistemic oughts, statements that seem to be sensitive to the agent's knowledge, and whose violations invite criticism of the agent. In addition, it briefly explores two directions for generalization: to relativistic oughts, and to conditional oughts. Even though the paper is long, its goals are modest. I have confined myself to the overall setting of stit semantics, and tried to show how epistemic oughts can be analyzed by generalizing the previous treatment of ordinary agentive oughts from [18] to the new framework of labeled

stit semantics—in particular, how epistemic oughts could be based on an ordering of action types from this new framework that mirrors the previous ordering on action tokens.

In following this narrow path, I have introduced several simplifications, which could be relaxed in a more substantial treatment. Two of these simplifications stand out as especially important. The first is the decision, from Section 4.1, to treat indistinguishability as a relation between moments, rather than as a relation between moment/history pairs, or indices. This decision gave rise, in Section 6.1, to the definition of information sets as sets of moments, rather than indices, and then in Section 7.2 to the restriction of our analysis of conditional oughts to statements with settled statements as antecedents. Because, in our central example, it is a settled fact at the time of your choice whether the coin has been placed heads up or tails up the analysis thus applies, as we have seen, to statements such as: If the coin is heads up, you ought to gamble. But because it is not yet settled whether or not you are going to win, the analysis does not apply to statements such as: If you are going to win, you ought to gamble. Treating indistinguishability as a relation between indices, not just moments, and so treating information sets as sets of indices, would allow us to address conditionals of this latter form, with contingent antecedents.

The second simplification I wish to highlight is the restriction set out right at the start, in Section 1.2, to stit models in which, at any moment, at most one agent faces a nontrivial choice. Relaxing this simplification would allow us, not only to analyze individual epistemic oughts in a richer setting, but to extend the present theory both to epistemic oughts involving groups of agents, and then to explore the relations between epistemic oughts bearing on groups and those bearing on the individual agents belonging to those groups. In the standard stit framework, without epistemic information, the relation of individual to group oughts—surely one of the most important matters in social philosophy—was addressed in a preliminary way in [18], and has recently received renewed and more sophisticated attention.[32] Moving the issue to an epistemic setting could be very rewarding.

[32]See, for example, Tamminga and Duijf [35] and Van De Putte [29].

References

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

[2] Alan Anderson. Logic, norms, and roles. *Ratio*, 4:32–49, 1962.

[3] Lennart Åqvist and Japp Hoepelman. Some theorems about a tree system of deontic tense logic. In Risto Hilpinen, editor, *New Studies in Deontic Logic*, pages 187–221. D. Reidel Publishing Company, 1981.

[4] Salvador Barberà, Walter Bossert, and Prasanta Pattanaik. Ranking sets of objects. In Salvador Barberà, Peter Hammond, and Christian Seidl, editors, *Handbook of Utility Theory, volume 2: Extensions*, pages 893–977. Springer Publishing Company, 2004.

[5] Nuel Belnap. Under Carnap's lamp: flat pre-semantics. *Studia Logica*, 80:1–28, 2005.

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

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

[8] Fabrizio Cariani, Magdalena Kaufmann, and Stefan Kaufmann. Deliberative modality under epistemic uncertainty. *Linguistics and Philosophy*, 2:225–259, 2013.

[9] Jennifer Carr. Subjective ought. *Ergo*, 2:678–710, 2015.

[10] Nate Charlow. What we know and what to do. *Synthese*, 190:2291–2323, 2013.

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

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

[13] Janice Dowell. Contextualist solutions to three puzzles about practical conditionals. In Russ Shafer-Landau, editor, *Oxford Studies in Metaethics*, pages 271–303. Oxford University Press, 2012.

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

[15] 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.

[16] Risto Hilpinen. On the semantics of personal directives. In C. H. Heidrich, editor, *Semantics and Communication*, pages 162–179. North-Holland Publishing Company, 1974.

[17] Risto Hilpinen and Paul McNamara. Deontic logic: a historical survey and introduction. In Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre, editors, *Handbook of Deontic Logic and Normative Systems*, pages 3–136. College Publications, 2014.

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

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

[20] John Horty and Eric Pacuit. Action types in stit semantics. *Review of Symbolic Logic*, 10:17–37, 2017.

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

[22] Stig Kanger. New foundations for ethical theory. Privately Distriubuted, 1957. Reprinted in Risto Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 36–58, D. Reidel Publishing Company, 1971.

[23] Niko Kolodny and John MacFarlane. Ifs and oughts. *Journal of Philosophy*, 107:115–143, 2010.

[24] 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.

[25] 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.

[26] John MacFarlane. *Assessment Sensitivity: Relative Truth and its Applications*. Oxford University Press, 2014.

[27] Derek Parfit. We we together do. Unpublished manuscript, 1988.

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

[29] Frederik Van De Putte. Choosing the right concept of right choice. Unpublished manuscript, 2018.

[30] Donald Regan. *Utilitarianism and Co-operation*. Clarendon Press, 1980.

[31] Leonard Savage. The theory of statistical decision. *Journal of the American Statistics Association*, 46:55–67, 1951.

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

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

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

[35] Allard Tamminga and Hein Duijf. Collective obligations, group plans, and individual actions. *Economics and Philosophy*, 33:187–214, 2017.

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

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

[38] Job van Eck. A system of temporally relative modal and deontic predicate logic and its philosophical applications. *Logique et Analyse*, 25:339–381, 1982.

[39] Malte Willer. A remark on iffy oughts. *Journal of Philosophy*, 109:449–461, 2012.