Deontic Epistemic *stit* Logic Distinguishing Modes of 'Mens Rea'

version as submitted to the Journal of Applied Logic

Jan Broersen

May 28, 2009

Abstract

Most juridical systems contain the principle that an act is only unlawful if the agent conducting the act has a 'guilty mind' ('mens rea'). Different law systems distinguish different modes of mens rea. For instance, American law distinguishes between 'knowingly' performing a criminal act, 'recklessness', 'strict liability', etc. I will show we can formalize several of these categories. The formalism I use is a complete *stit*-logic featuring operators for *stit*-actions taking effect in 'next' states, S5-knowledge operators and SDL-type obligation operators. The different modes of 'mens rea' correspond to the violation conditions of different types of obligation definable in the logic.

1 Introduction

An important distinction in law is the one between 'actus reus', which translates to 'guilty act', and 'mens rea' for 'guilty mind'. It is a general principle of law that both these conditions should be met for an act to qualify as criminal, that is, guilt not only presupposes a forbidden act as such, also, the performing agent must have committed the act knowingly, intentionally, purposely, etc.[1]. The task of showing that *both* necessary conditions 'actus reus' and 'mens rea' apply to an alleged criminal act, is in law referred to as 'showing concurrence'.

There are different levels of mens rea, each corresponding to different levels of culpability. And, of course, different law systems have different categories. The current North American system works with the following modes, in decreasing order of culpability (as taken from [18]):

- **Purposefully** - the actor has the "conscious object" of engaging in conduct and believes and hopes that the attendant circumstances exist.
- **Knowingly** - the actor is certain that his conduct will lead to the result.

[1] The general principle was already formulated back in 1797, by the English jurist Edward Coke: "actus non facit reum nisi mens sit rea", which is Latin for "an act does not make somebody guilty unless his/her mind is also guilty"

Dagstuhl Seminar Proceedings 09351
Information processing, rational belief change and social interaction
http://drops.dagstuhl.de/opus/volltexte/2009/2229

- **Recklessly** - the actor is aware that the attendant circumstances exist, but nevertheless engages in the conduct that a "law-abiding person" would have refrained from.
- **Negligently** - the actor is unaware of the attendant circumstances and the consequences of his conduct, but a "reasonable person" would have been aware
- **Strict liability** - the actor engaged in conduct and his mental state is irrelevant

The first class, the one of acts committed *purposefully*, is about acts that are instrumental in reaching an agent's malicious *goal*. The second class is not directly about an agent's intentions, aims or goals, but only about the condition whether or not an agent knows what it is doing. The third class is a little less clear. I think it is defendable to interpret it as the category of acts where an agent knowingly risks an unlawful outcome. For the fourth category, not knowing the (possible, or necessary - that is not made explicit) outcomes is not an excuse: if the agent did not know, it simply should have known. The final category concerns the complete absence of 'mens rea'. This is the category where agents can be culpable without having a 'guilty mind' whatsoever.

I claim the levels of culpability correspond to (1) levels of *excusability* and (2) levels of *deontic strength*. For the first class, the deontic strength is lowest of all and several excuses apply. In particular, for this class an 'actus reus' can be accompanied by the valid excuses: "I did not have bad intentions", "I did not know what I was doing", etc, etc. For the second category, deontic strength is higher, and less excuses apply. In particular, the excuse that there were no bad intentions is no longer acceptable. What counts is that the agent knew what it was doing, irrespective of the goal the act was aimed at. For the third category, where the deontic strength is yet higher, it is not even an excuse that the agent was not sure about the outcome: the agent is liable simply because it takes the *risk* the outcome is unlawful. In the fourth category, the excuse that the agent, "stupid enough", did not realize the consequences of his act, is no longer valid: for violations of any prohibition in this category he is still liable, because any 'reasonable' agent would have foreseen the consequences. And finally, for the strict liability category, deontic strength is highest of all, and no excuses referring to the mental state of an agent apply at all.

In philosophy, the idea that excuses play an important role in distinguishing different modes of acting was put forward by Austin [5]. And many other kinds of excuses than the ones above are thinkable. For instance, among the most well-known excuses for violating an obligation are: "I was not *able* to", "I do not agree my act *counts-as* a violation", "I obeyed a stronger, *conflicting* obligation" and "I did not *know* I had to". Of these, in this paper, I will only consider the first and the last one. The first one, about not being able to comply to the obligation, is only a valid excuse if the principle of "ought implies can" applies. The last one, concerning knowledge of the condition that the act is obliged, refers directly to the juridical principle "ignorantia juris non excusat",

which translates to "ignorance of or mistake about the law is no defence". So, here the (absence of) excuse is not so much about the mode of acting, as in the modes of mens rea above, but about whether or not the agent knows about the 'deontic status' of the act. This maybe subtle difference with the described modes of mens rea is not made very clear in the juridical literature. But, in our formalizations it will be.

We will also look at how we can formally define what counts as a 'reus actus'. Also for this, the juridical literature gives exact definitions. In particular, a reus actus cannot be an *involuntary* act. For instance, a person being thrown of a high building, surviving his fall by crashing into another person, who gets killed as the result of functioning as a cushion, has not committed an actus reus, even though the falling person knew that he actually was crashing into the person. The current American Model Penal Code [18] lists what acts count as involuntary acts for which no agent can be liable.

- a reflex or convulsion
- a bodily movement during unconsciousness or sleep
- conduct during hypnosis or resulting from hypnotic suggestion
- a bodily movement that otherwise is not a product of the effort or the determination of the actor, either conscious or habitual

In this paper we will formalize (1) the different modes of mens rea with the exception of the first category concerning 'purposefully' acts, (2) different modes of reus actus, that is, voluntary acts (3) the condition of "ignorantia juris non excusat". The mens rea class of 'purposefully' acts is not considered because I do not consider goals and intentions; I leave this for future research. Almost all the other categories concern conditions referring to an agent's *knowledge* about his actions. And knowledge operators will be a central concern of this paper. More in particular, we will come up with many different notions of obligation (as is common in deontic logic, we will treat obligations and prohibitions on a par, and see prohibitions as obligations to act oppositely), many of which can be associated with one of the classes of mens rea. The formal framework is also very well suited to refine and disambiguate the classes from the juridical literature.

This paper builds up the formal framework in three stages. First, in section 2, for the formalization of the acts as such, we define a *stit*-logic. Our *stit*-logic will be different from any *stit*-logic in the literature (with the exception of the earlier versions in [11] and [12]) in the sense that effects occur in 'next' states. In this section we make a start with our investigation of the modes of acting from the juridical literature by showing how to capture aspects of the 'voluntariness' requirement for an actus reus. Then, in section 3, we add an epistemic dimension to the logic, which will enable us to express the notion of 'knowingly doing'. This will enable us to be precise about the epistemic aspects of an actus reus. Then, third, in section 4 the deontic operators are introduced.

In this section we define the different types of obligation that correspond to different modes of mens rea.

2 A *stit*-logic affecting 'next' states: XSTIT

In this section we define a complete *stit*-logic where actions take effect in 'next' states: XSTIT. The logic XSTIT was first investigated in [11]. We also used the almost identical name 'X-STIT' in [13], but there the 'X' is separated from the acronym 'STIT', which refers to the fact that that paper's classical *instantaneous stit* logic is extended with a next operator, while in the present *stit*-variant effectivity of *stit*-operators itself refers to next states. That is not the only difference with the *stit*-logic(s) in [13]. In particular, the present logic drops some of the axioms in [13], adds several new ones, and is complete. Also we use a two dimensional semantics, closer to the *stit*-semantics in the philosophical literature.

The most distinguishing feature of the present *stit*-logic is that actions only take effect in 'next' states, where 'next' refers to immediate successors of the present state. This distinguishes the logic not only from the *stit*-variant in [13], but also from any *stit*-logic in the (philosophical) literature. However, there are very good reasons for taking this approach. The first reason is that the logics of the multi-agent versions of, what we might call, the standard 'instantaneous' *stit*, are undecidable and not finitely axiomatizable [6, 21]. The second reason is that the view that actions only take effect in some immediate next state, is the standard view in formal models of computation in computer science. And finally, also from a philosophical perspective, the choice can be advocated. Given that acting always seems associated with some effort or process, and given that these take time, we may conclude that actions take place 'in' time.

Besides the usual propositional connectives, the syntax of XSTIT comprises an operator $\Box\varphi$ for historical necessity, which plays the same role as the well-known path quantifiers in logics such as CTL and CTL* [19], and an operator $[A\ \mathsf{xstit}]\varphi$ for 'agents A jointly see to it that φ in the next state'. Given a countable set of propositions P and a finite set Ags of agent names, formally the language can be described as:

Definition 2.1 *Given a countable set of propositions P and $p \in P$, and given a finite set Ags of agent names, and $A \subseteq Ags$, the formal language $\mathcal{L}_{XSTIT}$ is:*

$$\varphi \quad := \quad p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [A\ \mathit{xstit}]\varphi$$

We define the temporal 'next' operator as the action performed by the complete set of agents Ags, leading to a unique follow-up state:

Definition 2.2

$$X\varphi \equiv_{def} [Ags\ \mathit{xstit}]\varphi$$

The view that the complete set of agents uniquely determines the next state is a common one. Not only it can be found in the multi-agent *stit*-logics in the philosophical literature [23], but also in related computer science formalisms such as Alternating Time Temporal Logic (ATL) [1, 2]. For the relation between *stit*-formalisms and ATL and Coalition Logic (CL) [28], see [14, 15].

Note that our *stit*-operator concerns, what game-theorists call, 'one-shot' actions. We can also imagine to have a *strategic stit*-operator (see [16]) where it is assumed that groups of agent have multiple subsequent choice points to ensure a certain condition (game theorist call these 'extensive games'). In my opinion, temporal operators other than the next operator make less sense for a *stit*-logic that is about one-shot actions. So, I am not in favor of, for instance, Horty's choice [23] to have a 'some time in the future' temporal operator in a non-strategic *stit*-setting. Ensuring a condition 'some time in the future' seems to me intrinsically strategic, and not often something that can be accomplished by a one-shot action. Interestingly enough, none of the one-shot *stit*-logics in the philosophical literature has a next operator, while from my point of view the next is actually the single most intuitive temporal operator to be considered in these logics.

In the description of the models, below, we will use terminology inspired by similar terminology from Coalition Logic, and call the relations interpreting the *stit*-operator 'effectivity' relations. However, our effectivity relations are *not* just the relational equivalent of the effectivity functions of CL. Our effectivity relations are relative to histories and determine the possible outcomes modulo the history. Effectivity functions in CL are relative to a state, and yield *sets* of possible outcomes.

Definition 2.3 *An XSTIT-frame is a tuple* $\langle S, H, R_\Box, \{R_A \mid A \subseteq Ags\}\rangle$ *such that:*

- S *is a non-empty set of states. Elements of* S *are denoted* s, s', *etc.*
- H *is a non-empty set of histories. Histories are sets of states. Elements of* H *are denoted* h, h', *etc.*
- *Structured worlds are tuples* $\langle s, h\rangle$, *with* $s \in S$ *and* $h \in H$ *and* $s \in h$.
- $R_\Box$ *is a 'historical necessity' relation over structured worlds such that* $\langle h, s\rangle R_\Box \langle h', s'\rangle$ *if and only if* $s = s'$
- *The* R_A *are 'effectivity' relations over structured worlds such that:*
  - R_{Ags} *is a 'next time' relation such that if* $\langle h, s\rangle R_{Ags} \langle h', s'\rangle$ *then* $h = h'$, *and* R_{Ags} *is serial and deterministic (the next state is completely determined by the choice made by the complete set of agents). So, histories 'contain' linearly ordered sets of states.*
  - $R_A \subseteq R_B$ *for* $B \subset A$ *(super-groups are at least as effective; in particular, effectivity for the empty 'group' and possibility for the complete group are inherited by all groups)*

- $R_\Box \circ R_{Ags} \subseteq R_\emptyset$ *('empty-group' effectivity implies system unavoidability / settledness)*
- $R_A \subseteq R_\Box \circ R_{Ags}$ *for any* A *(an action undertaken by* A *in the present state ensures the next state is element of a specific subset of all possible next states)*
- $R_{Ags} \circ R_\Box \subseteq R_A$ *for any* A *(no actions constitute a choice between histories that are undivided in next states)*
- *if* $\langle h,s\rangle R_\Box \langle h',s\rangle$ *and* $\langle h,s\rangle R_\Box \langle h'',s\rangle$ *then there is a* $\langle h,s\rangle R_\Box \langle h''',s\rangle$ *such that for* $A \cap B = \emptyset$*, if* $\langle h''',s\rangle R_A \langle h'''',s'\rangle$ *then* $\langle h',s\rangle R_A \langle h'''',s'\rangle$ *and if* $\langle h''',s\rangle R_B \langle h''''',s''\rangle$ *then* $\langle h'',s\rangle R_B \langle h''''',s''\rangle$ *(independence of group agency)*

The independence of agency condition above seem quite involved. The axiomatic correspondent in definition 2.6 is more clear. Mainly, what is expressed, is that possible effectivity sets of different agents have a non-empty intersection.

Definition 2.4 *A frame* $\mathcal{F} = \langle S, H, R_\Box, \{R_A \mid A \subseteq Ags\}\rangle$ *is extended to a model* $\mathcal{M} = \langle S, H, R_\Box, \{R_A \mid A \subseteq Ags\}, \pi\rangle$ *by adding a valuation* π *of atomic propositions:*

- π *is a valuation function* $\pi : P \longrightarrow 2^{S\times H}$ *assigning to each atomic proposition the set of history/state pairs in which they are true.*

The truth conditions for the semantics of the operators are standard. The non-standard aspect is the multi-dimensionality of the semantics. Note however that our logic is *not* a product logic [20], because the dimensions are not commutative.

Definition 2.5 *Validity* $\mathcal{M}, \langle s,h\rangle \models \varphi$*, of a formula* φ *in a history/state pair* $\langle s,h\rangle$ *of a model* $\mathcal{M} = \langle S, H, R_\Box, \{R_A \mid A \subseteq Ags\}, \pi\rangle$ *is defined as:*

$$
\begin{array}{lcl}
\mathcal{M}, \langle s,h\rangle \models p & \Leftrightarrow & \langle s,h\rangle \in \pi(p) \\
\mathcal{M}, \langle s,h\rangle \models \neg\varphi & \Leftrightarrow & \textit{not } \mathcal{M}, \langle s,h\rangle \models \varphi \\
\mathcal{M}, \langle s,h\rangle \models \varphi \wedge \psi & \Leftrightarrow & \mathcal{M}, \langle s,h\rangle \models \varphi \textit{ and } \mathcal{M}, \langle s,h\rangle \models \psi \\
\mathcal{M}, \langle s,h\rangle \models \Box\varphi & \Leftrightarrow & \langle s,h\rangle R_\Box \langle s',h'\rangle \textit{ implies that } \mathcal{M}, \langle s',h'\rangle \models \varphi \\
\mathcal{M}, \langle s,h\rangle \models [A\ \mathsf{xstit}]\varphi & \Leftrightarrow & \langle s,h\rangle R_A \langle s',h'\rangle \textit{ implies that } \mathcal{M}, \langle s',h'\rangle \models \varphi
\end{array}
$$

Satisfiability, validity on a frame and general validity are defined as usual.

While the semantics is standard from a (two-dimensional) modal logic perspective, the relation with standard *stit*-semantics deserves some explanation. In the conditions on the frames we recognize standard *stit* properties like 'no choice between undivided histories' and properties that are specific for the present *stit*-version, like 'actions take effect in successor states'. Actually, the frames can easily be pictured as trees where histories branch in states, like in standard *stit*-theory. The main difference is that *states* are not partitioned into choice sets. The choice sets appear here (implicitly) as sets of possible *next* states (like

in Coalition Logic). From a given 'actual' history/state pair, we reach these choice sets by first jumping (along $R_\Box$) to another history through the same state, and then looking (along R_A) what next states are reachable through the choice made by agents on that history.

One aspect of the present semantics needs extra clarification. Like in standard *stit*-semantics, history/state pairs for the same state can have different valuations of atomic propositions. In standard *stit*-formalisms this is actually needed to give semantics to the instantaneous effects of actions. But here, as said, the effects are not instantaneous. Therefore, in the present logic, the fact that different histories through the same state can have different valuations of non-temporal propositions, does not carry much meaning. Of course, in the logic we can talk about atomic propositions being true or not in other histories through the same state. For instance, the formula "$\Box p$" expresses that all the histories through the present state have in common that the atomic proposition p holds on them. But the point is that one might think that actually we should *impose* on the models that all histories through a state come with identical valuations of atomic propositions. That would induce the property $\varphi \rightarrow \Box\varphi$ for φ any '*stit*-operator-free' formula (in [13] we gave a system involving such an axiom). However, this would complicate establishing a completeness result, and does not strengthen the logic in any essential or interesting way. We think there is no need at all to impose such a condition. Since actions only take effect in next states, alternative valuations for *atomic* propositions on other histories through the same state are just not relevant for the semantics of the *stit*-fragment of our logic.

Now we go on to the axiomatization of the logic. Actually, axiomatization is fairly easy. The approach we have taken for constructing this logic is to build up the semantic conditions on frames and the corresponding axiom schemes simultaneously, while staying within the Sahlqvist class. This ensures that the semantics cannot give rise to more logical principles than can be proven from the axiomatization.

Definition 2.6 *The following axiom schemas, in combination with a standard axiomatization for propositional logic, and the standard rules (like necessitation) for the normal modal operators, define a Hilbert system for* XSTIT*:*

	S5 for $\Box$
	KD for each $[A\ \mathsf{xstit}]$
(Det)	$\neg X\neg\varphi \rightarrow X\varphi$
(C-Mon)	$[A\ \mathsf{xstit}]\varphi \rightarrow [A \cup B\ \mathsf{xstit}]\varphi$
($\emptyset \Rightarrow$ Sett)	$[\emptyset\ \mathsf{xstit}]\varphi \rightarrow \Box X\varphi$
(X-Eff)	$\Box X\varphi \rightarrow [A\ \mathsf{xstit}]\varphi$
(NCUH)	$[A\ \mathsf{xstit}]\varphi \rightarrow X\Box\varphi$
(Indep-G)	$\Diamond[A\ \mathsf{xstit}]\varphi \wedge \Diamond[B\ \mathsf{xstit}]\psi \rightarrow \Diamond([A\ \mathsf{xstit}]\varphi \wedge [B\ \mathsf{xstit}]\psi)$ *for* $A \cap B = \emptyset$

Theorem 2.1 *The Hilbert system of definition 2.6 is complete with respect to the semantics of definition 2.5.*

Proof The axioms correspond one-to-one to the semantic conditions defined on the frames (which can be easily verified with the on-line SQEMA system [17]), and are all within the Sahlqvist class. This means that all the axioms are expressible as first-order conditions on frames and that together they are complete with respect to the defined frame classes, cf. [9, Th.2.42]. ■

As part of the above axiomatization, we recognize Ming Xu's axiomatization for multi-agent *stit*-logics (see the article in [8]). Xu's axiomatization is for the standard, instantaneous *stit*-variant. The present *stit*-logic is different in two respects: (1) in the present logic, actions take effect in next states, and (2) the present logic is about groups of agents, while Xu's *stit* only considers individual agents. But, it should not come as a surprise that the same axioms apply to the present logic. The central property in Xu's axiomatization is the 'independence of agency' property. But the issue of independence of choices of different agents does not depend on the condition that effects are instantaneous or occur in next states.

Pauly's Coalition logic [28] is a logic of ability that is very closely related to *stit*-formalisms. In particular, in [14] it is shown that Coalition Logic can be embedded in *stit*-logic. Since in Coalition Logic actions also take effect in next states, restricting the *stit*-formalism by only allowing effects in next state, as in the logic of this paper, does not inhibit definability of Coalition Logic.

Theorem 2.2 *Coalition Logic, whose central operator is* $[A]\varphi$ *for 'agents* A *together can enforce* φ*', is embedded into the present logic by the definition* $[A]\varphi := \Diamond[A\ \mathsf{xstit}]\varphi$ *(plus the obvious isomorphic translations for other connectives).*

Proof Proof strategies similar to those in [14] and [13] can be applied. First we make sure that the axioms of coalition logic, after applying the above translation, are valid for the present logic. Applying the above translation to the CL-axioms from [28] yields:

($\bot$)	$\neg\Diamond[A\ \mathsf{xstit}]\bot$
($\top$)	$\Diamond[A\ \mathsf{xstit}]\top$
(N)	$\neg\Diamond([\emptyset\ \mathsf{xstit}]\neg\varphi \rightarrow \Diamond[Ags\ \mathsf{xstit}]\varphi$
(MON)	$\Diamond[A\ \mathsf{xstit}](\varphi \wedge \psi) \rightarrow \Diamond[A\ \mathsf{xstit}]\varphi$
(S)	$\Diamond[A\ \mathsf{xstit}]\varphi \wedge \Diamond[B\ \mathsf{xstit}]\psi \rightarrow \Diamond[A \cup B\ \mathsf{xstit}](\varphi \wedge \psi)$ for $A \cap B = \emptyset$

It is quite straightforward to verify these properties for the present logic, either semantically, or as theorems in the Hilbert system. ($\bot$) follows from KD for $[A\ \mathsf{xstit}]$. ($\top$) follows from normalilty of the operators. (N) follows from (X-Eff), the truth axiom for $\Box$, and propositional reasoning. (MON) follows from the normality of the operators. (S) follows from (Indep-G) and normality of $[A\ \mathsf{xstit}]$. Furthermore, we have to verify that the two coalition logic rules 'modus ponens' and 'logical equivalence' apply, but that is trivial, since both rules are sound a fortiori for the present normal modal system.

To complete the proof, we also have to show that the translation preserves validity in the other direction. Or, equivalently, we check that it preserves satisfiability in the same direction. That is, given that a CL formula is satisfiable on a CL-model, we have to show that its translation is satisfiable on the models defined in this paper. This is quite straightforward given the structural similarities between CL-models and the models in this paper.

■

Recall that the (N) axiom of CL corresponds to 'maximality' of Coalition Logic effectivity functions. Maximality of effectivity functions is the key property of so called 'playable' effectivity functions which Pauly proves to be equivalent to game forms. In the translation to XSTIT, as given above, we get $\neg\Diamond([\emptyset\ \mathsf{xstit}]\neg\varphi \rightarrow \Diamond[Ags\ \mathsf{xstit}]\varphi$. Now it turns out that this formula is not in the Sahlqvist class. We can also write it as $\Box\langle Ags\ \mathsf{xstit}\rangle\varphi \rightarrow \Diamond[\emptyset\ \mathsf{xstit}]\varphi$, where we recognize a variant on the well-known McKinsey property that is not first-order definable. But of course, it is very well possible that non-Sahlqvist axioms are derivable as theorems in a Sahlqvist logic.

Finally in this section, as a proposition we list some theorems.

Proposition 2.3 *The following are derivable in XSTIT:*

(1) $\Diamond[A\ \mathit{xstit}]\varphi \rightarrow \neg\Diamond([\overline{A}\ \mathit{xstit}]\neg\varphi$
(2) $\Diamond[\emptyset\ \mathit{xstit}]\varphi \leftrightarrow \neg\Diamond([Ags\ \mathit{xstit}]\neg\varphi$
(3) $[A\ \mathit{xstit}]\varphi \wedge [B\ \mathit{xstit}]\psi \rightarrow [A \cup B\ \mathit{xstit}](\varphi \wedge \psi)$
(4) $\Box X\varphi \rightarrow X\Box\varphi$
(5) $[A\ \mathit{xstit}]\varphi \rightarrow X\varphi$
(6) $X\varphi \leftrightarrow \neg X\neg\varphi$
(7) $\Box X\varphi \leftrightarrow [\emptyset\ \mathit{xstit}]\varphi$

Proof Derivation of all properties is just a little exercise in propositional normal modal logic. In the first property we recognize the 'regularity' property of Coalition Logic. It follows directly from (Indep-G).

■

Now, using XSTIT, we can already make a start with formalizing one of the concepts defined in the introduction. In particular, XSTIT will enable us to consider the notion of 'actus reus' more closely. As was explained, an actus reus must be a voluntary act. Some aspects of the concept 'voluntary' are captured by the *stit*-notion of 'deliberative action'. A deliberative *stit*-operator adds an extra condition to the standard XSTIT-operator, to avoid the property $[A\ \mathsf{xstit}]\top$. The idea is that agents should not be able to bring about things that will be true inevitably, but only things that without their intervention might not become true. We can easily define a deliberative version of the *stit*-operator, as follows:

$$[A\ \mathsf{dxstit}]\varphi \equiv_{def} [A\ \mathsf{xstit}]\varphi \wedge \neg\Box X\varphi$$

Theorem 2.4 *The operator $[A\ \mathit{dxstit}]\varphi$, is a minimal (i.e., weak) modal operator, not obeying weakening, or agglomeration, but obeying D.*

Proof The first part of the conjunction is KD and thus satisfies weakening, but the second part not, because of the negation. Because of the negation, the second part satisfies strengthening, but the first part not. The first part satisfies agglomeration, but the second part not. Both parts satisfy the D-axiom. ■

So, deliberateness, as defined in the operator above, seems to capture at least part of what it means to act voluntarily: one could also have acted otherwise, and thus one acts voluntarily. For instance, in the introduction, the crashing into the person breaking the fall of the man thrown of the building is not a voluntary act of the falling man, because the man had no choice but to fall, with the drastic consequence as a result. However, this is not the only thing we can say about voluntary / deliberate acts. But before we can go into this further, we will need to add an epistemic dimension to our *stit*-framework.

3 The concept of 'knowingly doing': E-XSTIT

In this section we extend XSTIT with epistemic operators $K_a\varphi$ for knowledge of individual agents a, resulting in the logic E-XSTIT. This will enable us to express the concept of 'knowingly doing'. Herzig and Troquard were the first to consider the addition of knowledge operators to a *stit*-logic [22]. Later on the framework was adapted and extended by Broersen, Herzig and Troquard [13, 16]. The E-XSTIT of the present section extends earlier versions in several ways. In particular, three axioms for the interaction of knowledge and action are proposed. Also the semantics, being two-dimensional, is different from the one in [13]. Finally, the modeled concept is 'knowingly doing', whereas in e.g. [22] the aim is to model 'knowing how'. In my opinion these concepts are different. I think 'knowing how' should be about whether an agent has a plan he knows to be effective. This to me seems an intrinsically strategic issue, one that cannot be approached in a non-strategic *stit*-setting. Also, 'knowing how' is an epistemic qualification concerning an *ability*, while 'knowingly doing' is an epistemic qualification concerning an *action*.

Definition 3.1 *Given a countable set of propositions P and $p \in P$, and given a finite set Ags of agent names, and $a \in Ags$ and $A \subseteq Ags$, the formal language $\mathcal{L}_{\text{E-XSTIT}}$ is:*

$$\varphi \ldots \quad := \quad p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_a\varphi \mid \Box\varphi \mid [A \text{ xstit}]\varphi$$

Definition 3.2 *An E-XSTIT-frame is a tuple $\langle S, H, R_\Box, \{R_A \mid A \subseteq Ags\}, \{\sim_a \mid a \in Ags\}\rangle$ such that:*

- *$\langle S, H, R_\Box, \{R_A \mid A \subseteq Ags\}\rangle$ is an XSTIT-frame*
- *The $\sim_a$ are epistemic equivalence relations over structured worlds such that:*

- $\sim_a \circ R_a \subseteq \sim_a \circ R_{Ags}$ *(agents cannot know what choices other agents perform concurrently)*
- $R_{Ags} \circ \sim_a \subseteq \sim_a \circ R_a$ *(agents recall the effects of the actions they knowingly perform themselves)*
- *if* $\langle h, s\rangle R_\Box \langle h', s'\rangle$ *and* $\langle h, s\rangle \sim_a \langle h'', s''\rangle$ *then there is a* $\langle h''', s'''\rangle$ *for which* $\langle h', s'\rangle R_\Box \langle h''', s'''\rangle$ *and if* $\langle h''', s'''\rangle R_a \langle h'''', s''''\rangle$ *then* $\langle h', s'\rangle (\sim_a \circ R_a) \langle h'''', s''''\rangle$ *(uniformity of conformant action)*

Definition 3.3 *Validity* $\mathcal{M}, \langle s, h\rangle \models \varphi$*, of a formula* φ *in a history/state pair* $\langle s, h\rangle$ *of a model* $\mathcal{M} = \langle S, H, R_\Box, \{R_A \mid A \subseteq Ags\}, \{\sim_a \mid a \in Ags\}, \pi\rangle$ *is defined as:*

All relevant clauses from definition 2.5, plus:

$$\mathcal{M}, \langle s, h\rangle \models K_a\varphi \quad \Leftrightarrow \quad \langle s, h\rangle \sim_a \langle s', h'\rangle \text{ implies that } \mathcal{M}, \langle s', h'\rangle \models \varphi$$

Satisfiability, validity on a frame and general validity are defined as usual.

With the above definitions we can express that agent a *knowingly* sees to it that φ as $K_a[a\ \mathsf{xstit}]\varphi$, where we slightly abuse notation by denoting $[\{a\}\ \mathsf{xstit}]\varphi$ as $[a\ \mathsf{xstit}]\varphi$. The semantics is in terms of models with epistemic equivalence sets (information sets) containing history/state pairs. An agent knowingly does something if its action 'holds' for all the history/state pairs in the epistemic equivalence set containing the *actual* history/state pair.

It is important to emphasize that the notion of 'knowingly doing' is entirely different from other notions combining knowledge and action or time in the literature. For instance, if we add epistemic uncertainty relations to temporal logic or dynamic logics, the choice is usually to define them over *states*. In that case uncertainty, and thus knowledge, cannot concern actions or choices themselves, but only state-determinate conditions. Only if we let uncertainty range over history/state pairs, as for the present logic, we can talk about (self-)knowledge of what agents are actually doing.

Definition 3.4 *The following axiom schemas, in combination with a standard axiomatization for propositional logic, and the standard rules (like necessitation) for the normal modal operators, define a Hilbert system:*

	All XSTIT *axioms determined by definition 2.6*
	S5 for each K_a
(Know-X)	$K_a X\varphi \to K_a[a\ \mathsf{xstit}]\varphi$
(Rec-Eff)	$K_a[a\ \mathsf{xstit}]\varphi \to XK_a\varphi$
(Unif-Str)	$\Diamond K_a[a\ \mathsf{xstit}]\varphi \to K_a\Diamond[a\ \mathsf{xstit}]\varphi$

Theorem 3.1 *The Hilbert system of definition 3.4 is complete with respect to the semantics of definition 3.3.*

Proof Like for XSTIT, the axioms for E-XSTIT correspond one-to-one to the semantic conditions, and are in the Sahlqvist class.

■

Before we elaborate on the interaction properties of E-XSTIT we list some properties that are derivable as theorems.

Proposition 3.2 *The following are theorems in E-XSTIT:*

$$K_a X\varphi \leftrightarrow K_a[a\ \mathit{xstit}]\varphi$$
$$K_a X\varphi \rightarrow X K_a\varphi$$

The last theorem in the list below is the well known 'perfect recall' or 'no forgetting' axiom, known from the literature on the interaction between epistemic and temporal modalities.

We now discuss each of the knowledge-action interaction properties in the E-XSTIT-semantics and axiomatization. The first one says that epistemic equivalence sets are closed under choices[2]. The corresponding axiom, is $K_a X\varphi \rightarrow K_a[a\ \mathsf{xstit}]\varphi$ (this property does not hold if the *stit*-operator is replaced by a *deliberative stit*-oparator). This property ensures that an agent cannot know that two histories belonging to the same choice are different, or, in other words, for any agent the histories within its own choices are indistinguishable. This means that agents cannot knowingly do *more* then what is affected by the choices they have. In particular, the property $K_a X\varphi \rightarrow K_a[a\ \mathsf{xstit}]\varphi$ says that agents can only know things about the (immediate) future if they are the result of an action they themselves knowingly perform. Then, an agent *unknowingly* does everything that is (1) true for all the history/state pairs belonging to the actual *choice* it makes in the actual *state*, but (2) not true for all the history/state pairs it considers possible. In general the things an agent does unknowingly vastly outnumber the things an agent *knows* it does. For instance, by sending an email, we may enforce many, many things we are not aware of, which are nevertheless the result of me sending the email. All these things we do *unknowingly* by knowingly sending the email.

Another, equivalent way of interpreting the property $K_a X\varphi \rightarrow K_a[a\ \mathsf{xstit}]\varphi$ is to say that it expresses that agents cannot know what actions other agents perform concurrently. This is because the independence property (Indep-G) guarantees that choices of other agents always refine the choices of the agent we consider. Then, knowing the choice of the other would mean that the agent would be able to know more about the future state of affairs then is guaranteed by his own action.

The second constraint on the interaction between knowledge and action is the one expressed by the axiom $K_a[a\ \mathsf{xstit}]\varphi \rightarrow X K_a\varphi$. The issue here is that if agents knowingly see to it that a condition holds in the next state, in that same next state they will recall that the condition holds. Like for the previous property, of course, I do not want to claim that this is a property that

[2] An extreme case is where the information sets are exactly the choices in each state. In that case an agent knows all the consequences of his actions.

is necessarily true for all systems of agents. Yet it is a property that we can impose for idealized agents that are not forgetful.

Finally, we discuss the interaction property $\Diamond K_a[a\ \mathsf{xstit}]\varphi \rightarrow K_a\Diamond[a\ \mathsf{xstit}]\varphi$. It says that if an agent can knowingly see to it that φ, then it knows that among its repertoire of choices there is one ensuring φ. This property is the *stit*-version of the constraint concerning 'uniform strategies' game theorists talk about. In game theory, *uniform* strategies require that agents have the same choices in all states within information sets. Since in game theory the choices are given names, a constraint is formulated saying that each state within the information set should have choices of the same type (that is, choices with the same name). In the present *stit*-setting, we do not have names. But the intuition that the same choices should be possible in different states of an information set, still applies. The property $\Diamond K_a[a\ \mathsf{xstit}]\varphi \rightarrow K_a\Diamond[a\ \mathsf{xstit}]\varphi$ exactly captures this intuition. It says that if an agent has the possibility to knowingly see to it that φ, then at least one of its choices in the states it considers possible actually ensures φ (that is, a φ-action is possible in all states of the information set). Maybe it is easier to see that the negation of the property, that is $\Diamond K_a[a\ \mathsf{xstit}]\varphi \wedge \widehat{K}_a\Box\langle a\ \mathsf{xstit}\rangle\neg\varphi$ (with $\widehat{K}_a$ the dual of K_a), is contradictory: it would be absurd if an agent has the possibility to knowingly see to it that φ and at the same time would consider it an epistemic possibility that it is settled that whatever it does, it allows for $\neg\varphi$ as a possible outcome. Yet another way of phrasing the property is to say that 'true ability' obeys the property of uniformity of strategies.

Now we can go back to formalizing the concepts discussed in the introduction. I explained that the standard deliberative *stit* captures part of the 'voluntariness' requirements for actus reus. However, voluntariness seems to involve more than just having had the possibility to do otherwise. Consider the following example. You carry a very dangerous contagious disease. But you do not know it. You travel by train and choose to sit next to some person and thereby unknowingly see to it that he is fatally infected. Now has an actus reus been committed (assuming spreading fatal diseases is forbidden by law)? The answer must be no. Even though it is true that you did spread the disease, and even though you could have done otherwise, what you did will not count as voluntarily or deliberately spreading the disease, simply because, to a certain extent, you did not know what you were doing.

So deliberateness or voluntariness entails both the possibility to do otherwise and having knowledge of what it is one is doing. Even more, an agent should have knowledge about the side-condition also: if an agent does not know that it could have done otherwise, we would not call the action deliberate. For the epistemic position on the side-condition, we then have two possibilities, motivating two new definitions for deliberate action:

$$[a\ \mathsf{dxstit}]'\varphi \equiv_{def} K_a[a\ \mathsf{xstit}]\varphi \wedge K_a\neg\Box X\varphi$$

$$[a\ \mathsf{dxstit}]''\varphi \equiv_{def} K_a[a\ \mathsf{xstit}]\varphi \wedge \neg K_a\Box X\varphi$$

The first notion says that deliberativeness requires that the agent not only knowingly performs the action, but also that the agent knows that the result is not settled, and thus that his action is needed to guarantee the result. The second notion has a different side-condition: the agent only considers it possible that the result is not settled.

Theorem 3.3 *The operators* $[a\ \mathsf{dxstit}]'\varphi$ *and* $[a\ \mathsf{dxstit}]''\varphi$ *are minimal (i.e., weak) modal operators, not obeying weakening, or agglomeration, but obeying D.*

Proof Considerations similar to those for theorem 2.4 apply. ■

By having suggested some definitions for capturing the voluntariness aspect of an actus reus, we have actually already touched upon the notion of mens rea. This is because talking about *epistemic* aspects of action clearly already introduces 'the mind' as a relevant concept in describing action. But we have not modeled any deontic aspects yet, and thus at this point we still cannot talk about the 'guilt' aspect of mens rea. Deontic aspects will be the subject of the next section.

4 Defining deontic modalities

For the extension of our framework with an operator for 'ought-to-do', we adapt the approach taken by Bartha [7] who introduces Anderson style ([3]) violation constants in *stit*-theory. The approach with violation constants is very well suited for theories of ought-to-do, witness the many logics based on adding violation constants to dynamic logic [25, 10]. However, we believe that the *stit*-setting is even more amenable to this approach. Some evidence for this is found in Bartha's article ([7]), that shows that many deontic logic puzzles (paradoxes) are representable in an intuitive way. And for the present paper a clear advantage of defining obligation as a reduction using violation constants, is that the completeness established for the logics in the previous sections is preserved after addition of the obligation operator.

Definition 4.1 *V is a violation constant $V \in P$.*

Bartha [7] defines his reduction for 'obligation to do' within the classical instantaneous *stit*-setting. Here we adapt that to the present situation where actions only take effect in next states. The intuition behind the definition is straightforward: an agent is obliged to do something if and only if by not performing the obliged action, it performs a violation. Since the effect of the obliged action can only be felt in next states, violations also have to be properties of next states. Formally, our definition is given by:

$$O[a\ \mathsf{xstit}]\varphi \equiv_{def} \Box(\neg[a\ \mathsf{xstit}]\varphi \rightarrow [a\ \mathsf{xstit}]V)$$

Theorem 4.1 *The operator $O[a\ \mathsf{xstit}]\varphi$ is KD, that is, it has the same properties as Standard Deontic Logic [30].*

Proof Rewrite $\Box(\neg[a\ \mathsf{xstit}]\varphi \rightarrow [a\ \mathsf{xstit}]V)$ as $\Box([a\ \mathsf{xstit}]\varphi \vee [a\ \mathsf{xstit}]V)$. Due to the propositional constant V, the part $[a\ \mathsf{xstit}]V$ is constant as a whole, which means that it does not affect the logical properties of the defined modal operator $O[a\ \mathsf{xstit}]\varphi$. The necessity operator $\Box$ is S5, and $[a\ \mathsf{xstit}]$ is KD. Using standard normal modal logic correspondence theory we conclude that the combined operator $\Box[a\ \mathsf{xstit}]\varphi$ is also KD.

■

The $\Box$ operator in the definition ensures that obligations are 'moment determinate'. This means that their validity only depends on the state, and not on the history (see [23] for a further explanation of this concept). We think that this is correct. But see [29] for an opposite opinion.

In this section we will not consider the 'side conditions' as in the previous sections. But these could, of course, easily be added to model the 'could have done otherwise' aspect of 'deliberateness'. Considering side-conditions would result in yet other categories.

Note that $\neg[a\ \mathsf{xstit}]\varphi$ expresses that a does not see to it that φ, which is the same as saying that a 'allows' a choice for which $\neg\varphi$ is a possible outcome. The definition then says that all such choices *do* guarantee that a violation occurs. So the agent is liable, because its action bore the risk of a bad outcome. The above defined obligation is thus a 'personal' one. If, by 'coincidence', φ occurs, apparently due the action of other agents, while the agent bearing the obligation did not make a choice that *ensured* that φ would occur, a violation is guaranteed. So agents do not escape an obligation by having other agents do the work for them.

We can also make the definition a little weaker and say that the agent is only liable if it actually guarantees the bad outcome:

$$O'[a\ \mathsf{xstit}]\varphi \equiv_{def} \Box([a\ \mathsf{xstit}]\neg\varphi \rightarrow [a\ \mathsf{xstit}]V)$$

Theorem 4.2 *The operator $O'[a\ \mathsf{xstit}]\varphi$ is a monotonic (i.e., weak) modal logic obeying the D axiom.*

Proof We have to check the properties of the combination $\Box\langle a\ \mathsf{xstit}\rangle\varphi$. We recognize a normal simulation of monotonic modal logic. Since S5 obeys D, the monotonic simulation inherits D.

■

Because the above two definitions do not at all refer to an agent's beliefs or other mental state, they both capture variants of the mens rea mode of 'strict liability'. For both definitions it is the case that if there is a violation, the agent is liable whatsoever, independent of whether or not the agent knows what it is doing. But, in my opinion this also includes the mens rea mode of 'negligently'. As described in the introduction, this class concerns those cases where 'a normal person' would have realized the consequences of his action. So, again, it does

not matter what that agent knows about what it is doing, it is liable whatsoever. The only difference with the 'strict liability' class is that there can be discussion about what a normal person can foresee, and thus, about whether something should be strictly liable or not.

Now we turn our attention to the mens rea classes 'knowingly' and 'recklessly'. It is clear that to define these, we can use the concept of 'knowingly doing' as defined in the previous section. We have several options, corresponding to different modes of mens rea. We discuss the following three modes:

$$OK[a\ \mathsf{xstit}]\varphi \equiv_{def} \Box(\neg K_a[a\ \mathsf{xstit}]\varphi \rightarrow [a\ \mathsf{xstit}]V)$$

$$OK'[a\ \mathsf{xstit}]\varphi \equiv_{def} \Box(K_a\neg[a\ \mathsf{xstit}]\varphi \rightarrow [a\ \mathsf{xstit}]V)$$

$$OK''[a\ \mathsf{xstit}]\varphi \equiv_{def} \Box(K_a[a\ \mathsf{xstit}]\neg\varphi \rightarrow [a\ \mathsf{xstit}]V)$$

The first operator, that is $OK[a\ \mathsf{xstit}]\varphi$, captures the mens rea mode of 'recklessly'. Here the agent has to knowingly see to it that φ obtains, since otherwise there will be a violation. In other words, if the agent is *reckless*, and does an action that it knows does not exclude an unlawful outcome, it is liable.

The third operator, that is $OK''[a\ \mathsf{xstit}]\varphi$, captures the mens rea mode of 'knowingly'. Here there is only a violation if the agent knowingly sees to it that the *opposite* of the lawful outcome φ obtains.

Finally, the second operator, that is $OK'[a\ \mathsf{xstit}]\varphi$ defines a mode of mens rea in between 'recklessly' and 'knowingly'. It says that the agent is liable if it knowingly refrains from obtaining φ. So, on the one hand, there is an aspect of recklessness: if the agent knowingly omits to do something, a violation occurs, because omitting may risk an undesirable consequence. On the other hand, if omitting is seen as a form of doing, we can also say that this expresses that there is a violation if the agent knowingly 'does' the for this level of mens rea inexcusable omission.

Theorem 4.3 *The operator $OK[a\ \mathsf{xstit}]\varphi$ is KD, that is, it has the same properties as Standard Deontic Logic [30]. The operators $OK'[a\ \mathsf{xstit}]\varphi$ and $OK''[a\ \mathsf{xstit}]\varphi$ are monotonic (weak) modal operator obeying the D axiom. In particular, the operators do not obey agglomeration.*

Proof For $OK[a\ \mathsf{xstit}]\varphi$ the proof is similar to the one for theorem 4.1. Here the knowledge modality is extra, which means that we have to investigate the logical behavior of the combination $\Box K_a[a\ \mathsf{xstit}]\varphi$, that is, a combination of S5, S5 and KD. This yields KD. For $OK'[a\ \mathsf{xstit}]\varphi$ and $OK''[a\ \mathsf{xstit}]\varphi$ the proofs are similar to the one for theorem 4.2

■

The final subject of this section is the principle of "ignorantia juris non excusat". For all of the above variants, nothing is said about whether or not the agent actually knows whether or not it has the obligation. But, we can associate awareness of an obligation directly with awareness of the act of bringing about

the violation in case of the agent not complying. So we can incorporate the principle by adapting the previous definitions as follows:

$$KOK[a\ \mathsf{xstit}]\varphi \equiv_{def} \Box(\neg K_a[a\ \mathsf{xstit}]\varphi \rightarrow K_a[a\ \mathsf{xstit}]V)$$

$$KOK'[a\ \mathsf{xstit}]\varphi \equiv_{def} \Box(K_a\neg[a\ \mathsf{xstit}]\varphi \rightarrow K_a[a\ \mathsf{xstit}]V)$$

$$KOK''[a\ \mathsf{xstit}]\varphi \equiv_{def} \Box(K_a[a\ \mathsf{xstit}]\neg\varphi \rightarrow K_a[a\ \mathsf{xstit}]V)$$

In these definitions also violations are knowingly brought about. This expresses that the agent bearing the obligation actually knows about the obligation, that is, the agent will knowingly bring about a violation if it does not comply with the obligation.

Theorem 4.4 *The operator $KOK[a\ \mathsf{xstit}]\varphi$ is KD, that is, it has the same properties as Standard Deontic Logic [30]. The operators $KOK'[a\ \mathsf{xstit}]\varphi$ and $KOK''[a\ \mathsf{xstit}]\varphi$ are monotonic (weak) modal operator obeying the D axiom. In particular, the operators do not obey agglomeration.*

Proof No difference with the properties for theorem 4.3 because the difference is only in the constant part of the operator definitions.

■

Of course, looking at the formal structure of the above definitions, a fourth definition suggests itself: one where it is not necessary to perform the obliged action knowingly, while at the same time, in case of non-compliance, the violation *is* brought about knowingly. But it seems clear right away that this combination is absurd. We cannot knowingly bring about a violation by unknowingly failing to comply with an obligation.

5 Related work

In [27] a logic is presented whose semantics shares several features with ours. In particular, the logic has epistemic indistinguishability relations ranging over history/state pairs. However, actions are omitted. In [26] actions are added to this framework by using action names in the models and the object language. So, the authors take a, what we might call 'dynamic logic view' on action. The work focusses on so called 'knowledge based obligations'. The central idea is that when agents get to know more, there are less histories they consider possible, which in turn may induce that the subset of deontically optimal histories, may give rise to new obligations. So the phenomenon being studied is that new knowledge may induce new obligations.

In our setting the phenomenon of getting more obligations by an increase in knowledge can occur in different ways. One way is simply by becoming aware of an obligation, that is, getting to know that one knowingly performs a violation by not performing some obliged action. Another route to enabling that obligations arise as the result of new knowledge, is by adopting the 'ought implies

can' principle for the stronger variants of our obligation operator. If agents get to know how to do something knowingly, they might incur an obligation that previously did not apply due to 'ought implies can'. This demonstrates that there seems to be more sides to the problem of 'knowledge based obligation'.

Another well-known interaction between epistemic and deontic modalities is Åqvist's puzzle of 'the knower' [4]. If knowledge is modeled using S5 and obligation using KD (SDL [30]), from $OK\varphi$ we derive $O\varphi$, which is clearly undesirable in an ought-to-be reading. However, this problem does not arise in the present logic, because obligation is strictly limited to apply to *actions*. In particular, if in Åqvist's example, for φ we substitute a *stit*-action $[\alpha\ \mathsf{xstit}]\varphi$, then we can read the derivation as 'the obligation to knowingly see to something implies the obligation to see to that same something'. In the present framework, that is not an undesirably property, but a desirable property obeyed by our definitions, because it is valid that $OK[a\ \mathsf{xstit}]\varphi \rightarrow O[a\ \mathsf{xstit}]\varphi$.

6 Future research

The logics we presented ask for extension in several ways. Note first that while the operators for agency are group operators, the operators for knowledge and obligation only refer to single agents. Actually, there are many open questions about how to generalize these operators to group operators. As is well-known, there are several notions of group-knowledge, such as 'shared knowledge', 'common knowledge' and 'distributed knowledge'. Which ones combine with which interaction properties for knowledge and group-action is yet unclear. Likewise we can consider generalizing the obligation operator to a group operator. Given the definitions of section 4 this actually hinges on providing group operators for the knowledge modalities.

Another issue concerns the violation constants. According to the present definitions, they are not relativized to agents or sets of agents. This corresponds to a 'consequentialist's' view on obligation, as in [23], where deontic optimality is determined according to an ordering of all possible histories. We could also take the view, like in [24], that deontic optimality orderings should be relative to agents or groups of agents. For our setting, using violation constants, that would mean that we introduce a violation constant for each agent or each group.

7 Conclusions

This paper presents an epistemic temporal *stit*-formalism that is complete with respect to a two-dimensional Kripke semantics. It introduces the new notion of 'knowingly doing' and discusses some of its possible properties. Using this notion, new 'epistemic' variants of operators for 'ought-to-do' are defined. In particular, many modes of 'mens rea' and characteristics of what counts as an 'actus reus', as defined in the juridical literature, can be analyzed and defined in the framework.

The first general conclusion to be drawn is that our logic framework is very useful for disambiguating and precisely defining action classes from the juridical literature. This is exemplified by the fact that in our definitions a new 'natural' level of mens rea in between 'knowingly' and 'recklessly' popped up. Furthermore, it is clear that I showed quite some restraint in defining different classes; many more subtle combinations are possible, for instance by demanding 'ought implies can', 'side conditions', etc. This suggests that the classification from the juridical literature could be much more subtle and fine-grained than it is, and the present framework could be of help in defining such a classification.

A second conclusion I want to draw is one about deontic logic in general. Sometimes, in discussions with other logicians, I have to defend deontic logic against the claim that there is not a single principle of deontic logic that is non-disputed. To a certain extent that is true. If one aims at designing 'the' logic of deontic reasoning, one is likely to end up with an extremely weak logic, since for every suggested principle, some deontic logician will raise his hand and come with a concrete scenario and the claim that this is a counter-example. In general, my claim would be that such counter-examples often introduce context that interferes with the pure deontic reasoning. For instance, the present paper makes very clear that the concept of action itself and the concept of knowledge may interact with the concept of obligation in many different subtle ways, giving rise to a whole plethora of definitions for ought-to-do. And then, action and knowledge are not even the only concepts interfering; there is also time, intention, etc. Then, what the present paper is also a clear example of is the phenomenon that if we want to account for all the modalities that interfere with the pure deontic modalities, and define deontic modalities acknowledging the interactions, we get weaker logics. And this mimics closely the complaint of logicians that there is not a single principle that is not disputed. My reply is thus, that the lack of logical properties is *not* inherent to deontic logic. It is only that deontic modalities often *appear* to be rather weak because they are contaminated with other, non-deontic modalities. And one of the tasks of deontic logicians is to expose the contamination, and bring all interfering modalities to the foreground. In particular, we can view the present work as part of a greater project in search for the 'building blocks' of deontic modalities. And, the building blocks investigated in this paper are 'action' and 'knowledge'.

References

[1] R. Alur, T.A. Henzinger, and O. Kupferman. Alternating-time temporal logic. In *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science*, Florida, October 1997.

[2] R. Alur, T.A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49(5):672–713, 2002.

[3] A.R. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 67:100–103, 1958.

[4] L. Åqvist. Good samaritans, contrary-to-duty imperatives, and epistemic obligations. *NOUS*, 1:361–379, 1967.

[5] J. Austin. A plea for excuses. *Proceedings of the Aristotelian Society*, (7), 1956.

[6] Philippe Balbiani, Olivier Gasquet, Andreas Herzig, François Schwarzentruber, and Nicolas Troquard. Coalition games over Kripke semantics: expressiveness and complexity. In Cédric Dègremont, Laurent Keiff, and Helge Rückert, editors, *Festschrift in Honour of Shahid Rahman*. College Publications, 2008. to appear.

[7] Paul Bartha. Conditional obligation, deontic paradoxes, and the logic of agency. *Annals of Mathematics and Artificial Intelligence*, 9(1-2):1–23, 1993.

[8] N. Belnap, M. Perloff, and M. Xu. *Facing the future: agents and choices in our indeterminist world*. Oxford, 2001.

[9] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2001.

[10] J.M. Broersen. *Modal Action Logics for Reasoning about Reactive Systems*. PhD thesis, Faculteit der Exacte Wetenschappen, Vrije Universiteit Amsterdam, februari 2003.

[11] J.M. Broersen. A complete STIT logic for knowledge and action, and some of its applications. In M. Baldoni, T. C. Son, M. B. van Riemsdijk, and M. Winikoff, editors, *Proceedings Workshop on Declarative Action Languages and Technologies (DALT) 2008*, Lecture Notes in Computer Science. Springer, 2008. to appear.

[12] J.M. Broersen. A logical analysis of the interaction between 'obligation-to-do' and 'knowingly doing'. In *Proceedings 9th International Workshop on Deontic Logic in Computer Science (DEON'08)*, volume 5076 of *Lecture Notes in Computer Science*, pages 140–154. Springer, 2008.

[13] J.M. Broersen, A. Herzig, and N. Troquard. A normal simulation of coalition logic and an epistemic extension. In *Proceedings Theoretical Aspects Rationality and Knowledge (TARK XI), Brussels*.

[14] J.M. Broersen, A. Herzig, and N. Troquard. From coalition logic to STIT. In *Proceedings LCMAS 2005*, volume 157 of *Electronic Notes in Theoretical Computer Science*, pages 23–35. Elsevier, 2005.

[15] J.M. Broersen, A. Herzig, and N. Troquard. Embedding Alternating-time Temporal Logic in strategic STIT logic of agency. *Journal of Logic and Computation*, 16(5):559–578, 2006.

[16] J.M. Broersen, A. Herzig, and N. Troquard. A STIT-extension of ATL. In Michael Fisher, editor, *Proceedings Tenth European Conference on Logics in Artificial Intelligence (JELIA'06)*, volume 4160 of *Lecture Notes in Artificial Intelligence*, pages 69–81. Springer Verlag, 2006.

[17] W. Conradie, V. Goranko, and D. Vakarelov. Algorithmic correspondence and completeness in modal logic I: The core algorithm SQEMA. *Logical Methods in Computer Science*, 2(1):1–26, 2006.

[18] M.D. Dubber. *Criminal Law: Model Penal Code*. Foundation Press, 2002.

[19] E.A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science, volume B: Formal Models and Semantics*, chapter 14, pages 996–1072. Elsevier Science, 1990.

[20] D.M. Gabbay, A. Kurucz, F. Wolter, and M. Zakharyachev. *Many-Dimensional Modal Logics: Theory and Applications*. Elsevier, 2003.

[21] Andreas Herzig and Francois Schwarzentruber. Properties of logics of individual and group agency. In Carlos Areces and Rob Goldblatt, editors, *Advances in Modal Logic*, volume 7, pages 133–149. College Publications, 2008.

[22] Andreas Herzig and Nicolas Troquard. Knowing How to Play: Uniform Choices in Logics of Agency. In Gerhard Weiss and Peter Stone, editors, *5th International Joint Conference on Autonomous Agents & Multi Agent Systems (AAMAS-06), Hakodate, Japan*, pages 209–216. ACM Press, 8-12 May 2006.

[23] J.F. Horty. *Agency and Deontic Logic*. Oxford University Press, 2001.

[24] Barteld Kooi and Allard Tamminga. Moral conflicts between groups of agents. *Journal of Philosophical Logic*, 37(1):1–21, 2008.

[25] J.-J.Ch. Meyer. A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29:109–136, 1988.

[26] E. Pacuit, R. Parikh, and E. Cogan. The logic of knowledge based obligation. *Knowledge, Rationality and Action a subjournal of Synthese*, 149(2):311–341, 2006.

[27] Rohit Parikh and Ramaswamy Ramanujam. A knowledge based semantics of messages. *Journal of Logic, Language and Information*, 12(4):453–467, 2003.

[28] Marc Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002.

[29] H. Wansing. Obligations, authorities, and history dependence. In H. Wansing, editor, *Essays on Non-classical Logic*, pages 247–258. World Scientific, 2001.

[30] G.H. von Wright. Deontic logic. *Mind*, 60:1–15, 1951.