Logic of Conditional Intention

Daniil Khaitovich[1*]

[1*]Institute for Logic, Language and Computation, University of Amsterdam, Science Park 107, Amsterdam, 1098 XG, The Netherlands.

Corresponding author(s). E-mail(s): d.khaitovich@uva.nl;

Abstract

This paper presents a modal logic of conditional intention. By conditional intention we understand a contingent commitment to an action that depends on the knowledge of whether means or reasons for the action obtain. The notion is studied in several philosophical [1, 2] and legal-theoretical papers [3–5], where it is shown to be crucial for practical reasoning. While several modal logics of unconditional intention were developed [6–10], the conditional intention is overlooked by modal logicians.

The paper contains both philosophical and mathematical-logical contributions to the problem. From the philosophical side, we show that several rationality norms of conditional belief and axioms for logics of conditionals should be rejected for the logic of conditional intention. From the technical side, we propose a new axiom system and prove that it is sound, complete, and has a finite model property with respect to a special class of Kripke models equipped with a set-selection function.

Keywords: Modal logic, Conditionals, Intention, Propositional attitudes, Mens Rea

1 Introduction

Since the foundational paper of [11], logicians are developing models of practical reasoning that incorporate intention as a separate, irreducible propositional attitude.[1] Philosophers of action and legal theorists lately emphasize the primary role of *conditional* intentions.[2] Broadly, conditional intention is a contingent commitment to a course of action, which depends on a certain condition. For example, if someone *unconditionally* intends to go for a walk, then they are committed to do so no matter what; but if they intend to go for a walk *in case of* good weather, then they are committed to do so as soon as they are sure that the

[1]See the introduction in [12] for a brief overview of logics of intention.
[2]See the introduction to [1] for an overview of conditional intention research.

weather is nice. Some philosophers argue that most real world intentions are conditional in nature: "*Often even an emphatic unconditional profession of intention such as 'I intend to φ at all costs' or 'I will φ no matter what' is not to be taken literally given that one is not really willing to φ at the price of losing one's life or making Heavens fall. Most intentions appear to be conditional in their 'deep structure' even when the conditions are not explicitly stated*" [2, p.700].

The neglect of conditional intention may have drastic consequences. In several earlier legal systems, *mens rea*[3] was often defined in terms of unconditional intention alone. This narrow formulation created a loophole: defendants could evade responsibility by claiming their actions were guided only by conditional intentions, which did not qualify as sufficient for *mens rea*.

> "*Irwin grabbed a woman visiting another prisoner and held a knife to her throat, threatening to kill her if he was not released. A jury convicted him of assault with intent to kill and he appealed on the grounds that he did not intend to kill the woman but intended only to kill her if his demands were not met. The Court of Appeals reduced his charge to one of assault without the requisite intention*" [4, p.275].[4]

Modal logics of propositional attitudes, including unconditional intention, are used to characterize *mens rea* in normative reasoning [14–16]. As the mentioned legal cases indicate, conditional intentions are crucial if we want to account for *mens rea*. This concern motivates the need to incorporate conditional intentions in the formal models of responsibility attribution. To our knowledge, no formal logic of conditional intention has been developed yet. We aim to fill in this gap.

The paper is organized as follows. In Section 2, we elaborate on our understanding of conditional intention, relying on the body of philosophical literature on the topic. In Section 3, we explicitly state the challenges our conceptual understanding of conditional intention poses for any formalization of this notion. Section 4 is dedicated to the formal exposition of our framework, resulting in Subsection 4.1, where a logic for our formal models is presented, which is sound, complete and has the finite model property. In Section 5, we list some optional refinements of our framework and argue that the presented logic is flexible enough to account for various philosophical positions about intention and practical reasoning. We conclude and list some directions for future work in Section 6.

2 Conditional intention: conceptual grounds

We adopt the general view on intention *simpliciter* à la Bratman [17]: intention is a commitment to a plan of action. Following [11], the view can be sloganized as "*intention is a choice with commitment*".[5] As for *conditional* intention, we rely on theories of [1, 2] that also inherit Bratmanian ideas. Conditional intention is a commitment to a contingent plan of action. When one says "*I intend to φ if ψ*"[6], one expresses a commitment to φ upon learning that ψ holds.

[3] *Mens rea* is a legal term meaning *"the mental element necessary to convict for any crime"[13].*

[4] Similar cases happened the U.S. and Great Britain, where analogous arguments were successfully used by defendants that were suspected of attempted theft, car hijacking, and some other crimes. A brief overview of these cases can be found in [3, 4].

[5] For the detailed overview of Bratman's views on intention in practical reasoning and their expansion in other fields, the reader may consult [6, 10, 17].

[6] Throughout the text, we use *one intends to φ* as an abbreviation for *one intends to bring about that φ*.

While searching for a suitable theory of conditional intention, one may try to reduce conditional intentions to the unconditional ones via material implication for the sake of parsimony. Unfortunately, such a reduction cannot be done. To clarify why it is impossible, we look into the next example borrowed from [2]. Assume that a nurse intends to change the patient's dressing in case he is still alive. Let us try to explain what the nurse's intention means using only unconditional intention. At first glance, we have at least two possible interpretations: *de re* and *de dicto*. *De re* interpretation is "*the nurse intends that if she knows that the patient is alive, then she changes his dressing*" ($Int(K\,alive \rightarrow change)$), *de dicto* is "*if the nurse knows that the patient is alive, then she intends to change his dressing*" ($K\,alive \rightarrow Int\,change$).[7] If we take the *de re* variant, then it suffices to falsify the antecedent of the intention to vacuously satisfy it. In this context, the nurse can kill the patient to make the implication $K\,alive \rightarrow change$ vacuously true, but it is not what she actually intends. In case of *de dicto*, if the bearer of the intention is not sure if the antecedent holds, we cannot derive anything about agent's intentions from the statement. In our example, as soon as the nurse is not sure if the patient is still alive, it does not follow that she has any intention, hence she might not be committed to any actions. On the contrary, it seems that even before the nurse gets to know that the patient is alive, she has a distinct motivational attitude: e.g., she needs to be prepared to change the patient's dressing. In other words, even if an agent is unaware whether the antecedent holds, the proposition about the conditional intention is informative: it tells us something about the motivation of this agent.

Therefore, neither *de re* nor *de dicto* interpretations account for the conditional intention in question, since by falsifying the antecedent of both formulations, we can trivialize the intention in undesirable ways. Since we need to distinguish between conditional intention and (necessary) material implications, from now on we will use different grammatical constructions for them. Namely, for conditional intention we will use *"one intends that* φ *in case of* ψ*"*, while *if-then* will be reserved for material implication.

More plausible interpretations of conditional intention move beyond *de re* and *de dicto* reductions. We adopt the terminology of [2] and call them *preconditional* and *restrictive conditional* intention:

- Preconditional intention: one preconditionally intends to φ in case of ψ iff ψ is a necessary means of bringing about that φ. Several examples: I intend to take a bus, in case the bus is coming; I intend to pay, in case I have money. In all such cases, one cannot do what is intended unless the condition of the intention holds. Obviously, it is not enough to falsify the precondition to satisfy the preconditional intention: the bearer of the intention usually wants its precondition to hold, since it is necessary to achieve what is intended;
- Restrictive conditional intention: one has a restrictive conditional intention to φ in case of ψ iff one is committed to φ only when one is sure that ψ obtains, otherwise, one lacks reasons to φ. In other words, the condition of one's intention restricts the scope of situations under which the agent is committed to a certain course of action. A number of examples: I intend to take a walk in case it does not rain; I intend to go to a library in case it is noisy at my office. In all the listed cases, the condition motivates one to fulfil the intention, while it is not a necessary means of doing so: one can go to a library even if it is perfectly quiet in one's office or go for a walk in the rain, but may not be motivated to do so. The nurse of our

[7] Here we use pseudo-language, where $Int\,\varphi$ stands for "*the nurse intends that* φ*"*, $K\varphi$ stands for "*the nurse knows that* φ" and $\rightarrow$ is interpreted as material implication.

example has a restrictive conditional intention as well. She intends to change the patient's dressing in case he is still alive: otherwise, she still can do that, but there is no reason to do so.

Note that those are not two rival interpretations of the same phenomena, but rather two closely related concepts: *conditional intention* is an umbrella term for both of them.

Following [1], we state that conditional intention can indeed be reduced[8] to an intention *simpliciter* in two cases. Namely, if a (pre)condition is intended or known. For example, if I intend to be at the train station on time with a precondition that I get up early, and I intend to get up early, then I simply intend to wake up early and be at the train station on time. The same applies to the restrictive conditional intention: going back to the nurse example, if she intends to assist the patient in case she is sure that he is alive, while having full control over his life and intending to save him, then she simply intends to save him and assist him after. As for known conditions, it seems natural to assume that there is no need to conditionalize our intentions on propositions that we know are true: "*If one knows that a condition obtains, it is not a contingency, but simply part of the background of planning. A condition is therefore a candidate for being a conditional intention's antecedent (in brief, can be an antecedent) only if one does not take oneself to know it obtains.*" [1, p.38].[9]

Taking it all together, we understand the conditional intention to φ in case of ψ as a contingent commitment of its bearer to φ in case she gets to know that ψ obtains. ψ may be a reason or a means for the agent to φ. The notion is equivalent neither to *de re* nor to *de dicto* interpretations via unconditional intention and material implication. In irreducibly conditional intention, the condition should neither be intended nor known.

3 Logical challenges

In this section, we derive logical challenges from our philosophical intuitions about conditional intention. This way, we state the plausibility criteria that the modal logic of conditional intention should satisfy.

First, we define a formal language we are to work with, so that it is easier to talk about logical properties we are interested in. Given some countable set of propositional variables $Var = \{p_1, p_2, \ldots\}$, we define the formal language $\mathscr{L}$ of conditional intention logic:

$$\mathscr{L} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \vee \varphi) \mid \Box\varphi \mid Int^{\varphi}\varphi$$

where $p \in Var$. The reading of negation, disjunction, and all the derived Boolean connectives and constants ($\bot$, $\top$, $\wedge$, $\rightarrow$, $\leftrightarrow$) is standard. $\Box\varphi$ reads as "*it is necessary that* φ" or *"*φ *is a priori true"*. Throughout the paper, we idealize the intender whose reasoning we are to model as *practically omniscient*. By practical omniscience, we mean that whenever φ holds in all practically possible scenarios – i.e., the agents and the environment cannot behave in a way that falsifies φ – the intender a priori knows that φ. Therefore, the reading of $\Box\varphi$ as "*it is necessary that* φ" and *"*φ *is a priori true"* are not conflicting. We take such an idealization

[8] By reducibility of a conditional intention to the unconditional one we mean bi-conditional of the form $Int^{\psi}\varphi \leftrightarrow Int\,\varphi$, where $Int\,\varphi$ denotes an unconditional intention to φ. We will clarify and formally define the grounds for such a reduction in the following sections.

[9] In [2], it is argued that one may have an intention conditional on a known proposition. For a discussion between the two authors on that topic, see [1, 18].

to assume away aspects of bounded theoretical rationality that are not crucial of our inquiry, but would significantly complicate the formal exposition of the problem.

$Int^{\psi}\varphi$ stands for *"the agent implicitly intends that φ in case of ψ"*. By implicit intention, we mean that ψ a priori follows from what the agent intends in case of φ. It might be either an intended consequence or just a side-effect. We address the issue of formally distinguishing side effects in Sect.5.2, and adopt this interpretation for now to introduce the more complicated concepts and structures stepwise. For now, every time we write "the agent intends that φ conditional on ψ", we mean the implicit intention, unless stated otherwise. We write simply $Int\varphi$ instead of $Int^{\top}\varphi$ and read it as "*the agent intends that φ*". Our language does not distinguish between preconditional and restrictive conditional intentions: both are formalized as $Int^{\psi}\varphi$.

We list some desirable and undesirable logical principles that follow from philosophical literature on (conditional) intention and our stances we have outlined in Section 2. According to a generally accepted rationality constraint [7, 8, 10, 17], agents' intentions should be consistent. Analogously, agents' conditional intentions should be consistent w.r.t. the condition, i.e. if one intends to φ in case of ψ, then φ and ψ should be mutually consistent [1, 2]:

$$\models Int^{\psi}\varphi \rightarrow \Diamond(\psi \wedge \varphi) \tag{1}$$

It is widely acknowledged that two intentions should also be consistent w.r.t. one another: if one intends that φ and ψ (under the same condition), then it should be possible that $\varphi \wedge \psi$ (under that condition) [11, 17]:

$$\models (Int^{\psi}\varphi_1 \wedge Int^{\psi}\varphi_2) \rightarrow \Diamond(\psi \wedge \varphi_1 \wedge \varphi_2) \tag{2}$$

Some authors, including Bratman [17], go even further and argue that intentions should not only be agglomerate*able*, but should in fact agglomerate:

$$\models (Int^{\varphi}\psi_1 \wedge Int^{\varphi}\psi_2) \rightarrow Int^{\varphi}(\psi_1 \wedge \psi_2) \tag{3}$$

The agglomeration principle for intention is debatable. For instance, [19] contends that the unqualified version of this principle compels agents to form "*one enormous compound intention*" [p. 517] to guide their actions, which may undermine their ability to reason about such intention, given their boundedness in cognitive resources. Nonetheless, there are ways to defend the agglomeration of intentions against such a criticism; see, for example, [20] for a detailed discussion. Since we interpret $Int^{\varphi}\psi$ as " *the agent **implicitly** intends that ψ conditional on φ*", then agglomeration is natural: the "enormous compound intention" is a logical consequence of what the agent intends, but may not be explicitly present in her motivational attitude. We will get back to the ways of avoiding agglomeration in explicit intentions in Section 5.

Notably, agglomeration fails in Cohen and Levesque's logic [11], given their assumptions about the interplay between intentions and time. For instance, one may intend to have a dinner and to have lunch later, but may not intend to have a dinner and lunch later, since she intends to have a dinner and lunch at different times of the day, not simultaneously. We will treat the interplay between intention and time in a significantly different way than Cohen and

Levesque, so this line of agglomeration's criticism is not essential for us (we draw the comparison between our treatment of intentions' temporality with Cohen and Levesque's in in detail in Sec. 5.1).

As we have pointed out in the previous section, if a condition of a conditional intention is known or intended, the conditional intention should be reducible to the unconditional one:

$$\models \Box\varphi \to (Int^{\varphi}\psi \to Int\,\psi) \tag{4}$$

$$\models Int\,\varphi \to (Int^{\varphi}\psi \to Int\,\psi) \tag{5}$$

Next, we move to the undesirable principles. Following insights from the previous section, the condition of a conditional intention should be either a reason or a means for an agent to commit to a certain action. Obviously, not every proposition expresses a reason or a means to act. Some propositions are completely irrelevant to the agent's practical reasoning; hence, they are not rationally pressured to form a contingent commitment for every proposition. Formally speaking, the following set of formulae should be satisfiable for some φ: $\{\neg Int^{\varphi}\psi \mid \psi \in \mathscr{L}\}$ (i.e., the agent does not intend anything conditional on φ). Therefore, the *success axiom*, which is commonly accepted in logics of conditional propositional attitudes including belief [21, 22], should be falsifiable in the logic of conditional intention:

$$\not\models Int^{\varphi}\varphi \tag{6}$$

For example, whether it is going to rain in Prague is irrelevant to the practical matters of Amsterdam's residents. So, the agent in Amsterdam is not rationally pressured to intend that it will rain in Prague in case it will actually happen: $\neg Int^{rain}rain$.

Next, we face the issue of *monotonicity*. By unrestricted monotonicity, we mean the following:

$$Int^{\varphi}\psi \to Int^{(\varphi\wedge\theta)}\psi \tag{Mon}$$

i.e., conditional intention is safe under logically stronger conditions. We reject unrestricted monotonicity, since rational agents can accept "*plan-B intentions*". For example, one may intend to go for a walk in case it is sunny and, as a plan-B, intend to stay home in case it is sunny but they have not gone for a walk for whatever reason. Since the second intention explicitly specifies what to do in case the first one has failed, the bearer of such intentions will never find themselves in a situation where mutually inconsistent conditions of both contingent commitments in question hold. Formally:

$$Int^{(sunny\wedge\neg walk)}home, Int^{sunny}walk \not\rightsquigarrow Int^{(sunny\wedge\neg walk)}walk$$

where $\rightsquigarrow$ is a symbol to denote an intuitive notion of entailment in the ordinary language. On the other hand, monotonicity seems plausible when we do not deal with plan-A and plan-B intentions. For instance, if someone deliberates about what to do when it is sunny or windy, and intends to stay home in case it is windy, it is reasonable to assume they also intend to stay home in case it is both windy and sunny:

$$Int^{windy}home \rightsquigarrow Int^{(sunny\wedge windy)}home$$

As these examples indicate, unrestricted monotonicity is too limiting because it prevents agents from having rational plan-B intentions.

$$Int^{\varphi}\psi \not\models Int^{(\varphi\wedge\vartheta)}\psi \tag{7}$$

Some weaker forms of monotonicity should also fail. If what is intended conditional on φ is consistent with the stronger condition $\varphi\wedge\vartheta$ – and even with the intentions under that stronger condition – it is not sufficient to ensure that the intention is safe under that stronger condition:

$$Int^{\varphi}\psi\wedge\Diamond(\varphi\wedge\vartheta\wedge\psi) \not\models Int^{(\varphi\wedge\vartheta)}\psi \tag{Mon1}$$

$$Int^{\varphi}\psi\wedge\neg Int^{(\varphi\wedge\vartheta)}\neg\psi \not\models Int^{(\varphi\wedge\vartheta)}\psi \tag{Mon2}$$

For instance, assume an agent plans to go for a run if the weather is good. As a part of that plan, she intends to wear running clothes in case of good weather. She also has a backup plan: if the weather is good, but she has failed to go for a run, she intends to stay home instead. Obviously, it is possible for her to wear running clothes even if she does not go running. Her backup plan does not specify what to wear, so she does not intend *not* to wear the running clothes in case she fails to go for a run on good weather. Still, the agent is not rationally pressured to intend to put on the sportswear in case the weather is good but she does not go for a run. To wear the running clothes is a part of the agent's primary plan – to go for a run conditionally on good weather – that happens to be consistent with her backup plan – to stay home conditionally on the good weather and the failure to go for a run. However, within the context of the backup plan, this part of the primary plan (to put on the sportswear) serves no purpose. Formally, let φ, ϑ, and ψ be formulae denoting "*the weather is good*", "*one is not going for a run*", and "*one puts on the running clothes*", respectively. Then, the story we have just presented constitutes a counterexample for the two weaker forms of monotonicity described above.

So to adequately capture the failure of monotonicity, we should track which intentions are parts of which plans. Given two conditions φ and $\varphi\wedge\vartheta$, if one intends that ψ conditional on φ, whether one is rationally pressured to intend that ψ conditionally on $\varphi\wedge\vartheta$ depends on whether $\varphi\wedge\vartheta$ contradicts the whole plan the agent has conditional on φ, not just the separate conditional intention that ψ.

We can look at a closely related logic in search of parallels: the logic of conditional obligations. In normative systems, we often face requirements of the form "*in case of* φ*, it is obligatory that* ψ". We constantly face *contrary-to-duty obligations* that specify what we are obliged to do in case we have failed to act upon other obligations [23]. This sounds very similar to plan-B intentions. Contrary-to-duty obligations cause the logics of conditional obligations to falsify monotonicity, but they are not the only reason. Consider the following example:

You ought not to act aggressively no matter what.	(A)
In case you meet your worst rival, you ought to act aggressively.	(B)

For A and B to be consistent, unrestricted monotonicity should fail. Such cases are sorted out via the principle *lex specialis derogat legi generali*: obligations with more specific conditions override those with less specific ones [24, 25]. Here, A has no condition whatsoever (or a trivial condition $\top$), while B has the condition, which is why in case of a conflict we should follow B, not A.

The difference between obligations and intentions in this regard is that the former may be exo- and heterogeneous. Obligations may come from different sources, some of which are external forces, and this causes collisions [26]. A number of examples: one may be obliged not to be rude due to general etiquette and ought to be rude to a specific person due to a custom in the given social context; one may have an obligation not to restrict anyone's freedom of speech, which yields from Universal Declaration of Human Rights, and ought to actively suppress certain kind of expressions due to domestic law of a certain country; one ought to avoid unnecessary spending due to economic rationality and ought to spend money on a specific unnecessary present as a sign of generosity due to a social norm. Unlike obligations, intentions are indo- and homogenous. They only come from the practical reasoning of the bearer of the intention, and no external force can put an intention in someone's mind [17]. Therefore, collisions, analogous A-B, can be avoided for intentions. Using the very same example, it seems irrational to simultaneously intend not to be aggressive to anyone and intend to be aggressive to one's rival. One's intention is only up to oneself, so even if one is confronted with two contradictory requirements from two different sources, one needs to choose which (if any) to follow and formulate the intention accordingly.

We take an intermediate position between unrestricted monotonicity, which is too strong, and *lex specialis derogat legi generali*, which is too weak. Namely, our principle is *antecedent-consequent consistency*. Given two conditions φ and $\varphi \wedge \vartheta$, as soon as **everything** the agent intends in case of φ is consistent with $\varphi \wedge \vartheta$, the agent should intend it in case of $\varphi \wedge \vartheta$ as well. In other words, if one has an intention conditional on φ, one must save that intention under a logically stronger condition $\varphi \wedge \vartheta$, as soon as that stronger condition does not contradict one's plan for what to do in case of φ. Such a solution still allows agents to have plan-A and plan-B intentions: the condition of plan-B contradicts what is intended in plan-A, therefore, our principle will not pressure agents to merge such plans into one contradictory plan.

To motivate our principle, we provide a number of examples. Let us go back to the following pair of statements, which expresses plan-A and plan-B intentions and intuitively should be consistent:

> In case it is sunny, Inne intends to go for a walk. (C)
>
> In case it is sunny, but Inne is not going for a walk, Inne intends to stay home. (D)

Obviously, from C it does not follow that Inne intends to go for a walk in case it is sunny and she is not going for a walk: the condition $\mathit{sunny} \wedge \neg \mathit{walk}$ contradicts her intention to go for a walk, so monotonicity may fail here. Therefore, C and D do not lead us into inconsistency. On the contrary, the following pair of statements should rightfully lead us to inconsistency:

> In case it is sunny, Inne intends to go for a walk. (E)
>
> In case it is windy, Inne intends to stay home. (F)

Intuitively, E-F should not be consistent. Let us assume that it is sunny *and* windy. Should Inne still follow E or F? In case it is sunny and windy, Inne still can go for a walk or stay home; nothing rationally pressures her to abandon any of her conditional commitments, but it is impossible to fulfill them both.

E-F are indeed inconsistent, following our monotonicity principle. Everything Inne intends in case it is sunny – to go for a walk – is still possible to do in case it is sunny and windy, therefore Inne must intend to go for a walk in case it is sunny and windy. Analogously, Inne is also required to intend to stay home in case it is sunny and windy, but it is impossible to simultaneously stay home and go for a walk. Therefore, if we adopt the proposed monotonicity principle, E-F together with 1 and 3 lead us to inconsistency, as we would intuitively expect.

To conclude this section, we have outlined two major challenges for the logic of conditional intention: cautious selection of conditions and restricted monotonicity. Both challenges pose specific requirements that are caused by our conceptual view on conditional intention, therefore, we cannot simply borrow solutions from "neighboring" logics, such as logics of conditional belief or conditional obligation. As for the selection of conditions, we will use the modification of formalisms proposed in [21]. Namely, as we are to see in the following section, we can make set-selection functions partial to avoid the conditionalization of intentions on any proposition whatsoever. As for restriction on monotonicity, we have argued for a principle that is stronger than *lex specialis derogat legi generali*: according to our principle, conditional intention should be monotonic, as soon as strengthening of the condition does not contradict the primary intention. In the following section, we will give a formally precise definition of this principle and show with examples why it works better than *lex specialis derogat legi generali*.

4 Formal semantics

Having pointed out the logical challenges and our views on them in Section 3, we present formal semantics for $\mathscr{L}$. We implement models with set-selection functions à la [21].

Definition 1 (**C**-frames and models) A class of conditional frames **C** consists of tuples of the form $F = (W, f)$, where:

1. $W \neq \emptyset$ is a set of practically possible worlds. By practical possibility, we mean that worlds of W are possible with respect to actions that the agent can perform. If a world is logically possible but would not be real no matter what the agent does, it is not practically possible. For example, whilst it is logically possible that the agent in question has never been born, it is not practically possible;
2. $f : (W \times 2^W) \to 2^W$ is a *partial* set selection function, which takes a possible world, a proposition $P \subseteq W$[10] and in case the agent conditionalizes her intentions on P in w, returns a set of conative P-alternatives of w, i.e. a set of possible worlds, where agent's intention in case of P is satisfied. f function has the following additional properties:
   (a) If $f(w,P)$ is defined, then $f(w,P) \cap P \neq \emptyset$: agent's intentions in case of P are consistent with P;

[10] Following Chellas, we will call sets of possible worlds $P \subseteq W$ *propositions* [21]. The reader should not confuse propositions in this sense with *propositional variables*, i.e. $p_1, p_2, \ldots \in Var$.

(b) If $f(w,P)$ and $f(w,Q)$ are defined, while $P \cap Q \neq \emptyset$, then $f(w,P \cap Q)$ is defined: if P and Q are mutually consistent reasons or means for agent's actions, then $P \cap Q$ is one as well;
(c) If $f(w,P)$ and $f(w,P \cap Q)$ are defined and $f(w,P) \cap P \cap Q \neq \emptyset$, then $f(w,P \cap Q) \subseteq f(w,P)$: in case everything the agent intends in case of P is possible to realize in case of $P \cap Q$, her $P \cap Q$-intention should include her P-intention;
(d) Given that $f(w,P)$ is defined, if $f(w,P) \subseteq Q$ and $f(w,P \cap Q)$ is defined, then $f(w,P) = f(w,P \cap Q)$: if agent intends that Q in case of P, then P-intention and $P \cap Q$-intention are the same.

As usual, every frame $F = (W,f) \in \mathbf{C}$ can be extended to a model $M = (F,V)$ with a valuation function $V : Var \to 2^W$.

Definition 2 (Semantic clauses for $\mathscr{L}$) Given a model $M = (W,f,V)$ and a possible world $w \in W$, where $[\![\varphi]\!]_M := \{w \in W \mid M,w \models \varphi\}$, the satisfiability relation $\models$ is inductively defined as follows:

$$
\begin{aligned}
&M,w \models p \Leftrightarrow w \in V(p)\\
&M,w \models \neg\varphi \Leftrightarrow M,w \not\models \varphi\\
&M,w \models \varphi \vee \psi \Leftrightarrow M,w \models \varphi \text{ or } M,w \models \psi\\
&M,w \models \Box\varphi \Leftrightarrow \forall v \in W : M,v \models \varphi\\
&M,w \models Int^{\varphi}\psi \Leftrightarrow f(w,[\![\varphi]\!]_M) \text{ is defined and } f(w,[\![\varphi]\!]_M) \subseteq [\![\psi]\!]_M
\end{aligned}
$$

$M \models \varphi$ means that $M,w \models \varphi$ for every $w \in W$. $F,w \models \varphi$ means that $(F,V),w \models \varphi$ for every valuation function V. $F \models \varphi$ means that $F,w \models \varphi$ for every $w \in W$. $\mathbf{C} \models \varphi$ means that $F \models \varphi$ for every $F \in \mathbf{C}$. When the class of frames is clear from the context, we write $\models \varphi$. In that case, we say that φ is valid.

We introduce a useful acronym: $[C]\varphi := Int^{\varphi}\top$. Intuitively, $[C]\varphi$ means that the agent conditionalizes her intentions on φ. Formally, $M,w \models [C]\varphi$ iff $f(w,[\![\varphi]\!]_M)$ is defined.

Although we are using the standard set-selection functions for conditionals, there are two critical differences with the semantics proposed in [21]. The first difference is our interpretation of the set W. Namely, we take W to be a set of *practically* rather than logically possible worlds. We take the agent to be reasoning about some practically feasible options she chooses, so she is free not to take into account some purely logical possibilities that are irrelevant to her practical matters. The second difference is that the set-selection function f is partial: for some proposition P and some possible world w, $f(w,P)$ may be undefined, which means that P is not a reason or a precondition for any action the agent deliberates over in w.

Despite the partiality of f, there are no truth-value gaps. From Def. 2 it follows that if $f(w,[\![\varphi]\!]_M)$ is undefined, then $M,w \models \neg Int^{\varphi}\psi$ for any ψ. If the agent does not conditionalize her intention on φ, then the agent does not intend anything at all conditional on φ. E.g., $f(w,[\![\varphi]\!]_M)$ is undefined iff $M,w \models \neg Int^{\varphi}\top$. This is an iff-statement, since if $f(w,[\![\varphi]\!]_M)$ is defined, then $f(w,[\![\varphi]\!]_M) \subseteq [\![\top]\!]_M = W$ by definition. This feature of partiality grounds our interpretation of the acronym $[C]\varphi$. If the agent has some intentions conditional on φ, then whatever is intended in case of φ, $\top$ is an a priori consequence of such an intention. $[C]\varphi := Int^{\varphi}\top$ says nothing informative about *what* is intended in case of φ, but expresses that at least something is intended in case of φ, i.e., φ is a condition.

Since there exists a pointed model (M,w), where the agent does not intend anything whatsoever conditional on φ, including tautologies, Int^{φ}-modality is not normal. Namely, the rule

of necessitation fails for Int^{φ}: from $\models \psi$ it does not follow that $\models Int^{\varphi}\psi$. For instance, take any pointed model (M,w), such that $f(w,V(p))$ is not defined. Then, $M,w \not\models Int^{p}\top$, while $\models \top$.

To demonstrate how the conditional models handle non-monotonicity, we proceed with an example.

Example 1 Imagine a self-driving car Dosen with two sensors: the first detects other cars around Dosen and the second detects pedestrians who happen to be on the road. Dosen unconditionally intends to avoid road accidents no matter what: $Int\neg\texttt{crash}$. It is the strongest unconditional intention of Dosen. If the first sensor reports that there are no cars in the traffic on the left side, then Dosen intends to turn to merge in: $Int^{\texttt{free}}\texttt{turn}$.

Dosen has the following true statements in its' knowledge base. If it tries to take a turn while there is a pedestrian on the road on the same side, it will inevitably lead to an accident: $\Box((\texttt{turn}\wedge\texttt{human})\rightarrow\texttt{crash})$. There may be no cars, but a pedestrian on the same side of the road: $\Diamond(\texttt{free}\wedge\texttt{human})$. It is possible to avoid the accident when there is a pedestrian and no cars on the same side: $\Diamond(\texttt{free}\wedge\texttt{human}\wedge\neg\texttt{crash})$.

Dosen tries to form an intention of what to do in case there are no cars but a pedestrian on the left side of the road. If he had adopted *lex specialis derogat legi generali*, then the more specific intention to merge in case there is a space free of cars would override the unconditional intention not to crash. In other words, if there was free space and a pedestrian on the road, Dosen would intend to merge even though it would lead to a road accident. Obviously, we do not want Dosen to reason in such fashion. Using our logic, can Dosen figure out what to do in case there are no cars on the left side, but there is also a pedestrian (i.e. $\texttt{free}\wedge\texttt{human}$)?

Assume a model $M=(W,f,V)$ for this case, where $w\in W$ is the real world. For the sake of brevity, we will not fully specify the whole model. Assume that Dosen's intentions are specified for three conditions: no matter what, i.e. $f(w,W)$, there are no cars on the left, i.e. $f(w,[\![\texttt{free}]\!])$, and there are no cars but a pedestrian, i.e. $f(w,[\![\texttt{free}\wedge\texttt{human}]\!])$. We know Dosen's strongest unconditional intention, to avoid crashes, i.e. $f(w,W)=[\![\neg\texttt{crash}]\!]$; In case of free space, Dosen intends to take a turn, i.e. $f(w,[\![\texttt{free}]\!])\subseteq[\![\texttt{turn}]\!]$. Since $f(w,[\![\texttt{free}\wedge\texttt{human}]\!])$ is defined, Dosen has specified some intention what to do in case there is a free space and a pedestrian on the left. We argue that this intention is consistent with Dosen's unconditional intention. As stated in Dosen's knowledge base, it is possible to avoid a crash while there is a pedestrian and free space to the left, i.e. $[\![\neg\texttt{crash}\wedge\texttt{free}\wedge\texttt{human}]\!]\neq\emptyset$. In other words, everything Dosen intends unconditionally is possible to realize in case there are no cars and a pedestrian on the left. Therefore, by 2(d) of Definition 1, $f(w,[\![\texttt{free}\wedge\texttt{human}]\!])\subseteq f(w,W)\subseteq[\![\neg\texttt{crash}]\!]$. Then, $f(w,[\![\texttt{free}\wedge\texttt{human}]\!])\subseteq[\![\neg\texttt{crash}]\!]$. By definition of $\models$, it follows that $M,w\models Int^{\texttt{free}\wedge\texttt{human}}\neg\texttt{crash}$.

We also argue that $\texttt{free}$-intention is not monotonic w.r.t. $\texttt{free}\wedge\texttt{human}$-intention, i.e. $f(w,[\![\texttt{free}\wedge\texttt{human}]\!])\not\subseteq f(w,[\![\texttt{free}]\!])$. To see it, recall that, as we have just shown, $f(w,[\![\texttt{free}\wedge\texttt{human}]\!])\subseteq[\![\neg\texttt{crash}]\!]$ and, as specified by example, $f(w,[\![\texttt{free}]\!])\subseteq[\![\texttt{turn}]\!]$. For the sake of contradiction, assume that Dosen intends to take a turn in case there is a free space and a pedestrian on the left, i.e. $f(w,[\![\texttt{free}\wedge\texttt{human}]\!])\subseteq[\![\texttt{turn}]\!]$. Since $f(w,[\![\texttt{free}\wedge\texttt{human}]\!])\subseteq[\![\neg\texttt{crash}]\!]$, it follows that $f(w,[\![\texttt{free}\wedge\texttt{human}]\!])\subseteq[\![\neg\texttt{crash}\wedge\texttt{turn}]\!]$. Then, by 2(a), it should be possible that $\texttt{free}\wedge\texttt{human}\wedge\texttt{turn}\wedge\neg\texttt{crash}$. As specified in Dosen's knowledge base, in case he turns while there is a human,

it will necessarily lead to a crash, i.e. $\Box((\texttt{human} \wedge \texttt{turn}) \to \texttt{crash})$. Therefore, $[\![\texttt{free} \wedge \texttt{human} \wedge \texttt{turn} \wedge \neg\texttt{crash}]\!] = \emptyset$, what contradicts 2(a) of Definition 1. To conclude, in case there is a free space and a pedestrian on the left of Dosen, he still intends to avoid crashes and therefore does not intend to take turns:

$$M, w \models Int^{(\texttt{free} \wedge \texttt{human})} \neg\texttt{crash} \wedge \neg Int^{(\texttt{free} \wedge \texttt{human})} \texttt{turn}$$

We proceed with showing some interesting (in)validities on conditional models. First, the set of conditions is restricted, since f function is partial. In other words, given any model $M = (W, f, V)$, $f(w, [\![\varphi]\!]_M)$ is not necessarily defined for any $w \in W$ and any $\varphi \in \mathscr{L}$, hence, $\not\models [C]\varphi$. By the same reason, success axiom (6) fails: $\not\models Int^{\varphi}\varphi$.

Moreover, the success axiom can be falsified even if the condition is defined: $\not\models [C]\varphi \to Int^{\varphi}\varphi$. This failure is warranted: even when φ is a condition, the agent is not rationally pressured to intend that φ in case of φ. The difference with the logic of conditional belief regarding the success axiom highlights that intention is a *motivational* rather than a *representational attitude*. Even though the agent is rationally pressured to believe that φ under the condition that φ, it does not justify that the agent should be (implicitly) motivated to bring about that φ in the same case. For instance, a driver may intend to turn back to her own lane in case she ends up in the oncoming lane. Despite conditionalizing her intentions on driving in the wrong lane, the driver is not rationally pressured to intend to drive in the wrong lane in case she does so. Formally, $[C]lane \not\models \neg Int^{lane} lane$, which reflects this informal intuition.

Our logic indeed validates the principles 1-5 we have informally argued for:

Proposition 1 *The following principles are valid in the class of* **C***-frames:*

$$\models Int^{\psi}\varphi \to \Diamond(\varphi \wedge \psi) \tag{8}$$

$$\models (Int^{\varphi}\psi_1 \wedge Int^{\varphi}\psi_2) \to Int^{\varphi}(\psi_1 \wedge \psi_2) \tag{9}$$

$$[C]\varphi \models \Box\varphi \to (Int\,\psi \leftrightarrow Int^{\varphi}\psi) \tag{10}$$

$$[C]\varphi \models Int\,\varphi \to (Int\,\psi \leftrightarrow Int^{\varphi}\psi) \tag{11}$$

Proof 1 holds, because of condition 2(a) of Definition 1; 3 holds, since if $f(w, [\![\varphi]\!]) \subseteq [\![\psi_1]\!]$ and $f(w, [\![\varphi]\!]) \subseteq [\![\psi_2]\!]$, then $f(w, [\![\varphi]\!]) \subseteq [\![\psi_1 \wedge \psi_2]\!]$. 4 holds, because if $M \models \Box\varphi$, then $[\![\varphi]\!]_M = [\![\top]\!]_M$, i.e. $f(w, [\![\varphi]\!]_M) = f(w, [\![\top]\!]_M)$. 5 holds, because of the condition 2(d) of Definition 1. □

Note that from (8) it follows that an agent does not conditionalize her intentions on a priori false propositions: $\models \Box\neg\varphi \to \neg Int^{\varphi}\psi$. From (10) it follows that if the proposition is a priori true, then any intention conditional on it is just an unconditional intention: $\models \Box\varphi \to (Int^{\varphi}\psi \leftrightarrow Int\,\psi)$. Therefore, our logic reflects Ludwig's thesis about conditions of intentions being "*as yet unsettled for us*" [1, p.32]: in a genuine conditional intention, the condition is neither a priori true, nor a priori false.

Third, we can deduce a restricted version of a *free choice*. The free choice is a common inference pattern, where conjunctive meanings are derived from disjunctive modal sentences, which goes against the prescriptions of classical logic [27]. We will demonstrate it with a deontic case. In the ordinary language, the expression "*one may smoke on the balcony* ***or***

outside" entails that one may smoke on the balcony **and** one may smoke outside (here, $P\varphi$ reads as "*it is permitted that* φ"):

$$P(\varphi \vee \psi) \rightsquigarrow (P\varphi \wedge P\psi) \tag{12}$$

It is not hard to find the analogous cases for conditional intention. For example, "*I intend to buy some apples in case they are tasty* ***or*** *cheap*" seems to entail that I intend to buy apples in case they are tasty **and** I intend to buy them in case they are cheap:

$$Int^{(\varphi \vee \psi)}\theta \rightsquigarrow (Int^{\varphi}\theta \wedge Int^{\psi}\theta)$$

In our logic, free choice inference is possible under two reasonable restrictions. First, such inference is possible if the agent conditionalizes her intention on $\varphi \vee \psi$, φ and ψ. Second, the inference is possible if there are no contradictions between φ, ψ and what is intended in case of $\varphi \vee \psi$. For example, assume that one strongly prefers walking by foot to driving a car. Therefore, in case one is either to drive or to walk, one intends to walk: $Int^{(walk \vee drive)}walk$. It is impossible to simultaneously drive and walk: $\neg\Diamond(walk \wedge drive)$. From this it seems implausible to infer that one intends to walk in case one is driving (and in our semantics, it indeed does not follow due to 2(a)):

$$\neg\Diamond(walk \wedge drive), Int^{(walk \vee drive)}walk \not\models Int^{drive}walk$$

In all other cases, i.e. when an agent has an intention conditional on disjunction $\varphi \vee \psi$ and both of the disjuncts, such that the intention in case of $\varphi \vee \psi$ is consistent with both φ and ψ, the free choice inference is valid.

Proposition 2 *The following principle is valid:*

$$\neg Int^{(\varphi \vee \psi)}\neg\psi \wedge \neg Int^{(\varphi \vee \psi)}\neg\varphi \wedge [C]\varphi \wedge [C]\psi \wedge [C](\varphi \vee \psi) \models (Int^{(\varphi \vee \psi)}\vartheta \rightarrow (Int^{\varphi}\vartheta \wedge Int^{\psi}\vartheta))$$

Proof See Theorem 3 of the following section. □

4.1 Logic, soundness, completeness

In what follows, we define an axiomatic system **L** (see Table 1 for the definition of **L**) and prove it to be sound and complete w.r.t. the class of **C**-models.

Axiom (C) ensures that if the agent intends anything conditional on φ, then φ should be a condition. (C-Inter) guarantees that whenever φ and ψ are consistent conditions, $\varphi \wedge \psi$ is a condition as well; this principle allows merging intentions with consistent conditions. Axiom (K) guarantees that intentions agglomerate and are closed under conjuncts, i.e. $Int^{\varphi}(\psi_1 \wedge \psi_2) \leftrightarrow (Int^{\varphi}\psi_1 \wedge Int^{\varphi}\psi_2)$. (U) ensures that $\Box$ behaves as a universal modality. (D+) ensures that conditional intention is consistent. (Cut) is a generalization of a schema $([C]\varphi \wedge Int\,\varphi) \rightarrow (Int\,\psi \leftrightarrow Int^{\varphi}\psi)$, i.e., if the condition of one intention is (conditionally) intended, such a

(CPL)	Tautologies of classical propositional logic
(S5)	S5 axioms and rules for $\Box$
(C)	$Int^{\varphi}\psi \rightarrow [C]\varphi$
(K)	$Int^{\varphi}(\psi \rightarrow \theta) \rightarrow (Int^{\varphi}\psi \rightarrow Int^{\varphi}\theta)$
(U)	$(\Box\psi \wedge [C]\varphi) \rightarrow Int^{\varphi}\psi$
(D+)	$Int^{\varphi}\psi \rightarrow \Diamond(\varphi \wedge \psi)$
(Cut)	$\big([C]\alpha \wedge [C](\alpha \wedge \psi)\big) \rightarrow \big(Int^{\alpha}\psi \rightarrow (Int^{\alpha}\varphi \leftrightarrow Int^{(\alpha\wedge\psi)}\varphi)\big)$
(R-Mon)	$([C]\alpha \wedge [C](\alpha \wedge \beta)) \rightarrow \big((Int^{\alpha}\varphi \wedge \neg Int^{\alpha}(\alpha \rightarrow \neg\beta)) \rightarrow Int^{(\alpha\wedge\beta)}\varphi\big)$
(C-inter)	$([C]\varphi \wedge [C]\psi \wedge \Diamond(\varphi \wedge \psi)) \rightarrow [C](\varphi \wedge \psi)$
(MP)	If $\vdash \varphi$ and $\vdash \varphi \rightarrow \psi$, then $\vdash \psi$
(RN$'$)	If $\vdash \psi$, then $\vdash [C]\varphi \rightarrow Int^{\varphi}\psi$
(RE)	If $\vdash \varphi \leftrightarrow \psi$, then $\vdash Int^{\varphi}\vartheta \leftrightarrow Int^{\psi}\vartheta$

Table 1 Logic **L**

conditional intention can be reduced.[11] (R-Mon) corresponds to the restricted monotonicity principle that we have argued for in Section 3.

Theorem 3 (Theorems and admissible rules of inference) *The following are derivable in* **L***:*

1. $Int^{\varphi}\psi \rightarrow \Diamond\varphi$
2. $Int^{\varphi}\psi \rightarrow \neg Int^{\varphi}\neg\psi$
3. *If* $\vdash \psi_1 \leftrightarrow \psi_2$, *then* $\vdash Int^{\varphi}\psi_1 \leftrightarrow Int^{\varphi}\psi_2$
4. *If* $\vdash \alpha \rightarrow \beta$, *then* $\vdash Int^{\varphi}\alpha \rightarrow Int^{\varphi}\beta$
5. *From* $\vdash ([C]\varphi \wedge [C]\psi \wedge [C](\varphi \vee \psi))$ *and* $\vdash (\neg Int^{(\varphi\vee\psi)}\neg\varphi \wedge \neg Int^{(\varphi\vee\psi)}\neg\psi)$ *it follows that* $\vdash (Int^{(\varphi\vee\psi)}\theta \rightarrow (Int^{\varphi}\theta \wedge Int^{\psi}\theta)$

Proof We derive only the last statement, since others are either easily derivable using D+ and K axioms or known to be derivable by standard methods in normal modal logics.[12]

1. $\neg Int^{(\varphi\vee\psi)}\neg\varphi \wedge \neg Int^{(\varphi\vee\psi)}\neg\psi$	
2. $Int^{(\varphi\vee\psi)}\theta$	Assumption
3. $[C]\varphi \wedge [C]\psi \wedge [C](\varphi \vee \psi)$	
4. $((\varphi \vee \psi) \rightarrow \neg\varphi) \leftrightarrow \neg\varphi$	CPL
5. $(Int^{(\varphi\vee\psi)}\theta \wedge \neg Int^{(\varphi\vee\psi)}((\varphi \vee \psi) \rightarrow \neg\varphi)) \rightarrow Int^{((\varphi\vee\psi)\wedge\psi)}\theta$	R-Mon, 3
6. $(\varphi \wedge (\varphi \vee \psi)) \leftrightarrow \varphi$	CPL $\Box$
7. $(Int^{(\varphi\vee\psi)}\theta \wedge \neg Int^{(\varphi\vee\psi)}\neg\psi) \rightarrow Int^{\psi}\theta$	From 3-5, RE
8. $Int^{\psi}\theta$	From 1,2, 6
9. $Int^{\varphi}\theta$	Analogously to 1-7
10. $Int^{\varphi}\theta \wedge Int^{\psi}\theta$	From 7,8
11. $Int^{(\varphi\vee\psi)}\theta \rightarrow (Int^{\varphi}\theta \wedge Int^{\psi}\theta)$	2-10, CPL

[11] We call this axiom schema (Cut) as a reference to the cut rule for the logic of conditionals: from $\vdash \alpha > \psi \wedge (\alpha \wedge \psi) > \varphi$ infer $\vdash \alpha > \varphi$, where $\varphi > \psi$ reads as "*conditional on* φ, ψ" [28]. If we replace $>$ with *Int*, we can rewrite it as an axiom in the following way: $Int^{\alpha}\psi \rightarrow (Int^{(\alpha\wedge\psi)}\varphi \rightarrow Int^{\alpha}\varphi)$, which is one of the directions of the bi-implication in the consequent of (Cut). The other direction – $Int^{\alpha}\psi \rightarrow (Int^{\alpha}\varphi \rightarrow Int^{(\alpha\wedge\psi)}\varphi)$ – follows from (R-Mon), since from $Int^{\alpha}\psi$, using (D+), it follows that $\neg Int^{\alpha}(\alpha \rightarrow \neg\psi)$. Hence, we could have written the (Cut) axiom in a weaker form, exactly as it is in the conditional logics, without loss of deductive power. We choose to write it in that redundant form just for the sake of intuitive understanding of what the axiom means.

[12] See different formulations of normal modal logics K and D in [29].

We move to proving the completeness of **L**. For any formula $\phi \in \mathscr{L}$, we will construct a special finitary language $\mathscr{L}_\phi$ and build a canonical model in a standard way, using maximal **L**-consistent subsets of $\mathscr{L}_\phi$ as possible worlds of the canonical model. As a consequence, the canonical model for every formula will be finite; hence, it follows that our logic has a finite model property.

Definition 3 (Restricted languages) Take any formula $\phi \in \mathscr{L}$. By a *modal depth* of ϕ, $md(\phi)$, we mean the deepest degree of nesting of modalities $\Box$ and Int in ϕ. In other words, $md(\phi)$ is recursively defined as follows:

$$md(p) = 0$$
$$md(\neg\phi) = md(\phi)$$
$$md(\phi_1 \vee \phi_2) = max(md(\phi_1), md(\phi_2))$$
$$md(\Box\phi) = 1 + md(\phi)$$
$$md(Int^{\phi_1}\phi_2) = 1 + max(md(\phi_1), md(\phi_2))$$

By $Var(\phi)$ we mean the set of propositional variables that occur in ϕ.

For every formula $\phi \in \mathscr{L}$, we define a special restricted language $\mathscr{L}_\phi$ as follows:

$$\mathscr{L}_\phi = \{\psi \in \mathscr{L} \mid md(\psi) \leq md(\phi), Var(\psi) \subseteq Var(\phi)\}$$

Given that (RE) rule is admissible for $\Box$ and Int modalities[13] and that there are finitely many logically non-equivalent modal-free formulae that contain finitely many propositional variables, $\mathscr{L}_\phi$ is *finite modulo* **L**[14] for any ϕ.

Definition 4 (MCSs and binary relations on them) Given any formula $\phi \in \mathscr{L}$, let $MCS_\phi \subseteq 2^{\mathscr{L}_\phi}$ be a collection of maximal **L**-consistent sets of formulae (shortly, mcs) of $\mathscr{L}_\phi$. Note that MCS_ϕ for any ϕ is finite: $\mathscr{L}_\phi$ is finite modulo **L** and for every $\Delta \in MCS_\phi$ and $\varphi \in \mathscr{L}_\phi$, $\varphi \in \Delta$ iff $\{\psi \in \mathscr{L}_\phi \mid \vdash_{\mathbf{L}} \varphi \leftrightarrow \psi\} \subseteq \Delta$.

Given any mcs $\Gamma \in MCS_\phi$, let $\Box\Gamma = \{\Box\psi \in \mathscr{L}_\phi \mid \Box\psi \in \Gamma\}$ be a $\Box$-theory of Γ, i.e. all the $\Box$-formulae that are in Γ. Then, $[\Gamma]_\Box = \{\Delta \in MCS_\phi \mid \Box\Gamma = \Box\Delta\}$ be a collection of maximal **L**-consistent sets of formulae that agree with Γ on all $\Box$-formulae. For any $\varphi \in \mathscr{L}_\phi$ and any mcs Γ, $Int^\varphi\Gamma = \{\psi \in \mathscr{L}_\phi \mid Int^\varphi\psi \in \Gamma\}$.

Definition 5 (Canonical model) Having some $\Gamma \in MCS_\phi$ fixed, we define a canonical model $M_\Gamma = (W_\Gamma, f_\Gamma, V_\Gamma)$, where:

- $W_\Gamma = [\Gamma]_\Box$;
- For any $\varphi \in \mathscr{L}_\phi$, $\widehat{\varphi} = \{\Delta \in W_\Gamma \mid \varphi \in \Delta\}$;
- $f_\Gamma(\Delta, \widehat{\varphi}) = \{\Sigma \in W_\Gamma \mid Int^\varphi\Delta \subseteq \Sigma\}$ if $[C]\varphi \in \Delta$. Otherwise, $f_\Gamma(\Delta, \widehat{\varphi})$ is undefined.
- $V_\Gamma(p) = \widehat{p}$ for any $p \in Var$.

First, we are to check that a canonical model is in **C**.

Lemma 1 (Any canonical frame is a conditional frame) For any $\Gamma \in MCS_\phi$, $(W_\Gamma, f_\Gamma) \in \mathbf{C}$.

---

[13] If $\vdash \varphi \leftrightarrow \psi$, then $\vdash \Box\varphi \leftrightarrow \psi$, $\vdash Int^\varphi\theta \leftrightarrow Int^\psi\theta$, $\vdash Int^\theta\varphi \leftrightarrow Int^\theta\psi$.

[14] A set of formulae S is finite modulo **L** iff there are finitely many classes of logically equivalent formulae in S. In other words, let $[\psi] = \{\varphi \in S \mid \vdash_{\mathbf{L}} \psi \leftrightarrow \varphi\}$. Then, S is finite modulo **L** iff the set $\{[\psi] \mid \psi \in S\}$ is finite.

Proof Since $\mathbf{L}$ is consistent, $W_\Gamma \neq \emptyset$ for any $\Gamma \in MCS_\phi$. It is a routine to check that $f_\Gamma : (W_\Gamma \times 2^{W_\Gamma}) \to 2^{W_\Gamma}$ is a well-defined partial function of the corresponding type.

2a: Assume that $f_\Gamma(\Delta, P)$ is defined for some $\Delta \in W_\Gamma, P \subseteq W_\Gamma$. Then, $P = \widehat{\varphi}$ for some $\varphi \in \mathscr{L}_\phi$ and $[C]\varphi \in \Delta$. For contradiction, assume that $f_\Gamma(\Delta, \widehat{\varphi}) \cap \widehat{\varphi} = \emptyset$. Then, there is no canonical world $\Sigma \in \widehat{\varphi}$, such that $Int^\varphi \Delta \subseteq \Sigma$. Then, $\{\varphi\} \cup \Box\Gamma \cup Int^\varphi \Delta \vdash \bot$. Then, (since $\Box\Gamma$ is consistent) there exists some $\psi_1, \ldots, \psi_n \in Int^\varphi \Delta$, such that $\Box\Gamma \vdash \neg(\varphi \wedge \psi_1 \wedge \ldots \wedge \psi_n)$. From S5 for $\Box$, $\Box\Gamma \vdash \neg\Diamond(\varphi \wedge \psi_1 \wedge \ldots \wedge \psi_n)$. Then, since $\psi_1, \ldots, \psi_n \in Int^\varphi \Delta$, $Int^\varphi \psi_1, \ldots, Int^\varphi \psi_n \in \Delta$. By (K) and (RN'), $Int^\varphi(\psi_1 \wedge \ldots \wedge \psi_n) \in \Delta$. Since $\neg\Diamond(\varphi \wedge \psi_1, \ldots, \psi_n) \in \Box\Gamma \subseteq \Delta$, $\neg\Diamond(\varphi \wedge \psi_1, \ldots, \psi_n) \in \Delta$. Hence, $Int^\varphi(\psi_1 \wedge \ldots \wedge \psi_n) \wedge \neg\Diamond(\varphi \wedge \psi_1 \wedge \ldots \wedge \psi_n) \in \Delta$, which contradicts (D+) axiom, which cannot happen, given that Δ by definition is $\mathbf{L}$-consistent. Therefore, $f_\Gamma(\Delta, P) \cap P \neq \emptyset$ for any Δ and P, s.t. $f_\Gamma(\Delta, P)$ is defined.

2b: If $f_\Gamma(\Delta, P)$ and $f_\Gamma(\Delta, Q)$ are defined, while $P \cap Q \neq \emptyset$, then $f_\Gamma(P \cap Q)$ is defined. Follows directly from (C-Inter) axiom.

2c: If $f_\Gamma(\Delta, P)$ and $f_\Gamma(\Delta, P \cap Q)$ are defined and $f_\Gamma(\Delta, P) \cap P \cap Q \neq \emptyset$, then $f_\Gamma(\Delta, P \cap Q) \subseteq f_\Gamma(\Delta, P)$. Assume that $f_\Gamma(\Delta, P)$ and $f_\Gamma(\Delta, P \cap Q)$ are defined. Then, $P = \widehat{\varphi_1}$, $P \cap Q = \widehat{\varphi_2}$ for some φ_1, φ_2, such that $\widehat{\varphi_2} \subseteq \widehat{\varphi_1}$ (i.e. $\widehat{\varphi_2} = \widehat{\varphi_1 \wedge \varphi_2}$), and $[C]\varphi_1, [C]\varphi_2 \in \Delta$. Assume that $f_\Gamma(\Delta, \widehat{\varphi_1}) \cap \widehat{\varphi_1 \wedge \varphi_2} \neq \emptyset$. It means that there exists some $\Sigma \in W_\Gamma$, such that $Int^{\varphi_1}\Delta \cup \{\varphi_1 \wedge \varphi_2\} \subseteq \Sigma$. We claim that $\neg Int^{\varphi_1}(\varphi_1 \to \neg(\varphi_1 \wedge \varphi_2)) \in \Delta$. For contradiction, assume otherwise. Then, since Δ is mcs, $Int^{\varphi_1}(\varphi_1 \to \neg(\varphi_1 \wedge \varphi_2)) \in \Delta$. Then, by definition of f_Γ, if $\Sigma \in f_\Gamma(\Delta, \widehat{\varphi_1})$, then $\varphi_1 \to \neg(\varphi_1 \wedge \varphi_2) \in \Sigma$. Take any $\Sigma \in f_\Gamma(\Delta, \widehat{\varphi_1}) \cap \widehat{\varphi_1 \wedge \varphi_2}$. Then, by our assumptions, $\varphi_1 \to \neg(\varphi_1 \wedge \varphi_2), \varphi_1 \wedge \varphi_2 \in \Sigma$. Then, $\Sigma \vdash \bot$, which contradicts the fact that Σ is mcs. Then, $f_\Gamma(\Delta, \widehat{\varphi_1}) \cap \widehat{\varphi_1 \wedge \varphi_2} = \emptyset$, which contradicts our initial assumption. Hence, $\neg Int^{\varphi_1}(\varphi_1 \to \neg(\varphi_1 \wedge \varphi_2)) \in \Delta$. Then, given that $[C]\varphi_1, [C](\varphi_1 \wedge \varphi_2) \in \Delta$, by (R-Mon) axiom we get that $Int^{\varphi_1}\psi \to Int^{(\varphi_1 \wedge \varphi_2)}\psi \in \Delta$ for any $\psi \in \mathscr{L}_\phi$. Then, $Int^{\varphi_1}\Delta \subseteq Int^{(\varphi_1 \wedge \varphi_2)}\Delta$, i.e. $f_\Gamma(\Delta, \widehat{\varphi_1 \wedge \varphi_2}) \subseteq f_\Gamma(\Delta, \widehat{\varphi_1})$, i.e. $f_\Gamma(\Delta, P \cap Q) \subseteq f_\Gamma(\Delta, P)$ (given that $P \cap Q = \widehat{\varphi_2} \subseteq \widehat{\varphi_1} = P$).

2d: If $f_\Gamma(\Delta, P) \subseteq Q$ and $f_\Gamma(\Delta, P \cap Q)$ is defined, then $f_\Gamma(\Delta, P) = f_\Gamma(\Delta, P \cap Q)$. Again, assume that $f_\Gamma(\Delta, P), f_\Gamma(\Delta, P \cap Q)$ are defined and $f_\Gamma(\Delta, P) \subseteq Q$. Then, $P = \widehat{\varphi}$, $P \cap Q = \widehat{\varphi \wedge \psi}$ for some $\varphi, \psi \in \mathscr{L}_\phi$, s.t. $[C]\varphi, [C](\varphi \wedge \psi) \in \Delta$, where $P = \widehat{\varphi}, Q = \widehat{\psi}$.[15]

By our assumption, $f_\Gamma(\Delta, \widehat{\varphi}) \subseteq \widehat{\psi}$. In other words, if $Int^\varphi \Delta \subseteq \Sigma$, then $\psi \in \Sigma$ for any $\Sigma \in W_\Gamma$. We claim that $Int^\varphi \psi \in \Delta$. Assume otherwise. Then, since Δ is mcs, $\neg Int^\varphi \psi \in \Delta$. The set $\Box\Gamma \cup Int^\varphi \Delta \cup \{\neg\psi\}$ is consistent. To prove this consistency claim, assume otherwise. Then, $\Box\Gamma \cup Int^\varphi \Delta \vdash \psi$. In other words, there exists $\vartheta_1, \ldots, \vartheta_n \in Int^\varphi \Delta$, such that $\Box\Gamma \cup \{\vartheta_1, \ldots, \vartheta_n\} \vdash \psi$. I.e., $Int^\varphi \vartheta_1 \wedge \ldots \wedge Int^\varphi \vartheta_n \in \Delta$ and $\Box\Gamma \vdash (\vartheta_1 \wedge \ldots \wedge \vartheta_n) \to \psi$. Then, by (K), $Int^\varphi(\vartheta_1 \wedge \ldots \wedge \vartheta_n) \in \Delta$ and $\Box\Gamma \vdash (\vartheta_1 \wedge \ldots \wedge \vartheta_n) \to \psi$. Then, since $\Box\Gamma \subseteq \Delta$, by the derivable rule $\vdash (\vartheta_1 \wedge \ldots \wedge \vartheta_n) \to \psi \Rightarrow \vdash Int^\varphi(\vartheta_1 \wedge \ldots \wedge \vartheta_n) \to Int^\varphi \psi$, $Int^\varphi \psi \in \Delta$, contradicting the fact that $\neg Int^\varphi \psi \in \Delta$ and Δ in consistent. Hence, $\Box\Gamma \cup Int^\varphi \Delta \cup \{\neg\psi\}$ is consistent. Then, there exists some mcs $\Sigma \in W_\Gamma$, such that $\Box\Gamma \cup Int^\varphi \Delta \cup \{\neg\psi\} \subseteq \Sigma$. Then, since $Int^\varphi \Delta \subseteq \Sigma$, $\Sigma \in f_\Gamma(\Delta, \widehat{\varphi})$. But $\neg\psi \in \Sigma$, while by our initial assumption $f_\Gamma(\Delta, \widehat{\varphi}) \subseteq \widehat{\psi}$. Contradiction.

Then, $Int^\varphi \psi \in \Delta$, from which it follows that $[C]\varphi \in \Delta$. Then, by (Cut) axiom, $[C](\varphi \wedge \psi) \to (Int^\varphi \theta \leftrightarrow Int^{(\varphi \wedge \psi)}\theta) \in \Delta$. Since $[C](\varphi \wedge \psi) \in \Delta$, $Int^\varphi \theta \leftrightarrow Int^{(\varphi \wedge \psi)}\theta \in \Delta$ for any $\theta \in \mathscr{L}_\phi$, from which it follows that $f_\Gamma(\Delta, \widehat{\varphi}) = f_\Gamma(\Delta, \widehat{\varphi \wedge \psi})$. □

Lemma 2 (Existence) For any $\Delta \in W_\Gamma$, if $\neg Int^\varphi \psi \in \Delta$, then there exists $\Sigma \in f_\Gamma(\Delta, \widehat{\varphi})$, s.t. $\neg\psi \in \Sigma$.

Proof Assume that $\neg Int^\varphi \psi \in \Delta$. Then, as we have shown in Lemma 1, $\Box\Gamma \cup Int^\varphi \Delta \cup \{\neg\psi\}$ is consistent. In this case, there exists a mcs $\Sigma \in W_\Gamma$, such that $\Box\Gamma \cup Int^\varphi \Delta \cup \{\neg\psi\} \subseteq \Sigma$. Since $Int^\varphi \Delta \subseteq \Sigma$, $\Sigma \in f_\Gamma(\Delta, \widehat{\varphi})$ and $\neg\psi \in \Sigma$. □

[15] We are sure that such φ, ψ exist, since 1) $W_\Gamma \subseteq MCS_\phi$ is finite and 2) every $\Delta \in W_\Gamma$ contains finitely many logically non-equivalent formulae. Hence, every canonical world Δ has its corresponding unique finite set of logically non-equivalent formulae S_Δ. Then, any set $P \subseteq W_\Gamma$ corresponds to $\bigvee_{\Delta \in P} \widehat{\bigwedge_{\psi \in S_\Delta} \psi}$.

Lemma 3 (Truth lemma) Given any $M_\Gamma = (W_\Gamma, f_\Gamma, V_\Gamma)$, $\Delta \in W_\Gamma$ and $\varphi \in \mathscr{L}_\phi$, $M_\Gamma, \Delta \models \varphi$ iff $\varphi \in \Delta$.

Proof By induction on the complexity of φ. Cases of φ being atomic or modal-free are trivial. Let $\varphi := \Box\psi$.

($\Rightarrow$) Assume $M_\Gamma, \Delta \models \Box\psi$. Then, $W_\Gamma = [\![\psi]\!]_{M_\Gamma}$, i.e. by IH $W_\Gamma = \widehat{\psi}$. Then, $\Box\Gamma \vdash \psi$, from which it follows that $\Box\psi \in \Delta$, since $\Box\Gamma = \Box\Delta$.

($\Leftarrow$) Assume $\Box\psi \in \Delta$. Then, since $\Box\Delta = \Box\Gamma = \Box\Sigma$ for any $\Sigma \in W_\Gamma$, $\widehat{\Box\psi} = W_\Gamma$. Then, by axiom T for $\Box$, $\widehat{\psi} = W_\Gamma$, i.e. by IH $[\![\psi]\!]_{M_\Gamma} = W_\Gamma$, i.e. $M_\Gamma, \Delta \models \Box\psi$.

Let $\varphi := Int^\alpha \psi$.

($\Rightarrow$) Assume $M_\Gamma, \Delta \models Int^\alpha \psi$. Then, $f_\Gamma(\Delta, [\![\alpha]\!]_{M_\Gamma}) \subseteq [\![\psi]\!]_{M_\Gamma}$. By IH, $f_\Gamma(\Delta, \widehat{\alpha}) \subseteq \widehat{\psi}$. For contradiction, assume that $Int^\alpha \psi \notin \Delta$. Then, $\neg Int^\alpha \psi \in \Delta$. By Lemma 2, there exists $\Sigma \in f_\Gamma(\Delta, \widehat{\alpha})$, such that $\neg\psi \in \Sigma$, which contradicts that $f_\Gamma(\Delta, \widehat{\alpha}) \subseteq \widehat{\psi}$. Hence, $Int^\alpha \psi \in \Delta$.

($\Leftarrow$) Assume that $Int^\alpha \psi \in \Delta$. Then, $f_\Gamma(\Delta, \widehat{\alpha}) \subseteq \widehat{\psi}$. By IH, $f_\Gamma(\Delta, [\![\alpha]\!]_{M_\Gamma}) \subseteq [\![\psi]\!]_{M_\Gamma}$, i.e. $M_\Gamma, \Delta \models Int^\alpha \psi$. □

Theorem 4 (Soundness and completeness) *Given any formula $\varphi \in \mathscr{L}$, the following holds:*

$$\mathbf{C} \models \varphi \Leftrightarrow \mathbf{L} \vdash \varphi$$

Proof ($\Leftarrow$) By the routine check of axioms' validity.

($\Rightarrow$) By contraposition. Assume that $\mathbf{L} \nvdash \varphi$. Then, $\{\neg\varphi\}$ is $\mathbf{L}$-consistent. Take a language $\mathscr{L}_{\neg\varphi}$, pick a mcs $\Gamma \ni \neg\varphi$ and build a model $M_\Gamma = (W_\Gamma, f_\Gamma, V_\Gamma)$. Then, $\neg\varphi \in \Gamma \in W_\Gamma$, from which by Lemma 3 it follows that $M_\Gamma, \Gamma \models \neg\varphi$. By Lemma 1, $(W_\Gamma, f_\Gamma) \in \mathbf{C}$, hence, $\mathbf{C} \not\models \varphi$. □

5 Time, action and and side-effects

So far, we have developed a formalism that captures key logical properties of conditional intention: restricted monotonicity and free choice inferences together with a limited set of conditions. In this section, we briefly discuss ways to refine our analysis with respect to other aspects: temporality, actions and the side-effects problem. We have not addressed these aspects from the start, since they may be redundant depending on the goals of the logic's user analysis and the varying philosophical stances she might hold.

5.1 Explicit time and actions

So far, we have implicitly treated the object of intention as a proposition about outcomes of an action the agent ruminates to do. Such a view presupposes a certain connection with actions and time. E.g., one cannot intend to change the past or to see to it that the future will be in a certain way if it is out of one's control. In this subsection, we will make the temporal structure and the agent's actions explicit, both in the language and semantics.

We analyze conditional intentions as *present-directed*: by intending that ψ conditional on φ, the agent contingently commits herself to a currently available action that brings about that ψ in case of φ. We do not deal with the *future-directed* conditional intentions, by which agents contingently commit themselves to some distant future action. For example, the author's

intention to finish typing this sentence in case their laptop keeps working is in the scope of our inquiry, while the future parent's intention to pay for their children education in case they will have children is not.

We concentrate on present-directed intention because of our interest in the connection between conditional intention and *mens rea*. Present-directed intentions are intentions *with which* agents act, and exactly present-directed intentions are important when we want to understand the mental component of the agents' responsibility for their actions. Future-directed intentions, on the other hand, may or may not bind agents to any actions whatsoever at the moment of deliberation. As Cohen and Levesque concisely put it, "[future-directed intentions] *guide agents' planning and constrain their adoption of other intentions, whereas* [present-directed intentions] *function causally in producing behavior*" [11, p.216].

For instance, it is important to understand which present-directed intentions Irvin had, when he grabbed a woman and held a knife to her throat. Did he intend to kill her conditional on not being released? Or did he intend to make an empty threat? It seems irrelevant whether he had malicious or merciful future-directed intentions towards what to do when and if he is released, since those did not bind him to assaulting the woman at the moment (at least not directly, but via the dependent present-directed intention).

Present-directedness distinguishes our analysis from the one of Cohen and Levesque. In their seminal work, they explicitly state that they analyze future-directed intention and every time they use the word "intention", they mean the future-directed one [11, p.216]. Therefore, a number of Cohen and Levesque's core issues – e.g., persistence of intention over time – are irrelevant to us. Yet the principles that are non-specific to future-directed intention are to be satisfied: rational present-directed intentions should still be consistent; a rational intender should (believe to) be able to satisfy her intention; and rational agents may not intend all the expected side-effects of their intention (to be discussed in Sec. 5.2).

Having clarified the essential relation between intention, actions, and time that we have in mind, we continue with the refinement of our formalism to make these points explicit. First, we extend the language $\mathscr{L}$ to $\mathscr{L}_t$ in the following manner:

$$\mathscr{L}_t \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \vee \varphi) \mid \Box\varphi \mid X\varphi \mid \varphi U \varphi \mid [a]\varphi \mid Int^{\varphi} X\varphi$$

where $X\varphi$ reads as *in the next moment in time, φ will hold*. $\varphi U \psi$ reads as *ψ is true now or will be true at some moment in the future, and φ holds until ψ*. We will use the widely accepted acronyms: $F\varphi := X(\top U \varphi)$ (*in some moment in future, φ holds*); $G\varphi := \neg F \neg \varphi$ (*from now on, it will always be true that φ*).

$[a]\varphi$ reads as *the agent sees to it that φ holds*. By seeing to it that we understand the behaviour which guarantees φ, regardless of all other agents and the environment. $\Box\varphi$ reads as *at the given moment, it is practically necessary that φ*. We still have the same understanding of practical necessity: φ is practically necessary iff φ is true independently of how all agents and the environment behave. We use standard acronyms from STIT (sees-to-it-that) logics: $\langle a \rangle \varphi := \neg[a]\neg\varphi$ (*a* behaves in a way that does not exclude φ); Standardly, $\Diamond[a]\varphi$ can be read as *a has an ability to see to it that φ* [30].

$Int^{\varphi} X\psi$ reads as *the agent intends that* ***in the next moment in time*** *ψ will hold, conditional on φ*. We adopt the acronym from Section 4: $[C]\varphi := Int^{\varphi} X \top$. We explicitly take the object of one's intention to be a proposition about the immediate consequences of one's currently available actions she commits herself to. Our language does not allow meaningless

expressions about intentions towards propositions in the past (e.g., *I intend that I have finished writing this subsection yesterday*) or in the moment of deliberation (e.g., *I intend that this subsection is already written now, as I am writing it*). No matter what the agent is going to do, propositions about the past or present are either true or false independently of her or anyone's choice of action; so propositions about the past and present are either practically necessary (if they are true) or practically impossible (if they are false).

We still can express the intention towards the long-term consequences of one's current actions: e.g., $Int^{\varphi}X(F\psi)$ reads as the agent intends that ψ will hold eventually after the next moment, conditional on φ. Yet this statement should not be confused with the expression of the future-directed intention! $Int^{\varphi}X(F\psi)$ means that the agent intends to do some currently available action, such that, conditional on φ, the action has *as its immediate consequence* that ψ will eventually be true. It does not mean that the agent intends to do some action in the future that guarantees ψ conditional on φ. For instance, I might have a present-directed intention that my colleague will eventually die from poisoning, conditional on her drinking her coffee. This intention causes me to poison her coffee. Given that my colleague drinks the poisoned coffee, she will survive the next couple of hours or even days, but will inevitably die sooner or later. Her inevitable death in the future is an immediate consequence of the action I am deliberating upon doing now. On the contrary, I may have a future-directed intention to poison my colleague. In that case, I still intend that my colleague will eventually die from poisoning, but my future-directed intention does not necessarily cause any action of mine at the given moment: I might commit to the act of poisoning in, e.g., the next year.

As for the semantics, instead of defining a space of practically possible worlds, we will work with branching time structures, where multiple futures are possible, depending on the course of agent's and the environment's behaviour. We choose to work with branching time, since it is the most natural framework to model agents' choices of action, where alternative courses of events are possible. For the motivation and philosophical background behind agency in branching time environments, see [30].

Definition 6 (**BT** structures) A branching time (in short, **BT**) structure is a tuple of the form $(T,<)$, where $T \neq \emptyset$ is a non-empty set of moments and $<\subseteq T \times T$ is an irreflexive, transitive and backward-linear[16] relation on T, ordering moments by precedence/succession. A **BT**-structure is discrete iff for any moment $m_n \in T$, there exists another moment $m_{n+1} \in T$, such that $m_n < m_{n+1}$ and there is no moment $m_{n'}$, such that $m_n < m_{n'} < m_{n+1}$.

Histories are maximal $<$-ordered chains of moments: $h \subseteq T$ is a history iff for any $m \in h$, if $m' < m$ or $m < m'$, then $m' \in h$ and for any $m, m' \in h$, either $m < m'$ or $m' < m$ or $m = m'$.

There are several alternative ways to model branching time. One of them is *concurrent game structures* (CGSs). From now on, we will work with CGSs, since it will allow us to deal with finite models (whereas **BT**-structures by definition are infinite). The only additional assumption we will have to take is *discreteness* of **BT**-structures. This assumption does not result in any conceptual consequences pertinent to our work.

Definition 7 (Concurrent game structures **CGS**) A concurrent game structure is a tuple $M = (W, Act, \{R_\delta\}_{\delta \in Act}, V)$, where:

[16] By backward linearity we mean that for any $m_1, m_2, m_3 \in T$, if $m_3 < m_1$ and $m_2 < m_1$, then either $m_2 < m_3$ or $m_3 < m_2$ or $m_3 = m_2$.

1. $W \neq \emptyset$ is a set of states;
2. $Act \neq \emptyset$ is a set of the agent's actions;
3. For each $\delta \in Act$, $R_\delta \subseteq W \times W$ is a binary relations on W, such that for every state $w \in W$, $\bigcup_{\delta \in Act} R_\delta(w) \neq \emptyset$.
4. $V : Var \to 2^W$ is a standard evaluation function.

Informally, for any two states w, v and any action $\delta \in Act$, $wR_\delta v$ means that doing δ at w may result in transition to v. Consequently, $R_\delta(w)$ is the set of possible outcomes of doing δ at w. If $R_\delta(w) = \emptyset$, then δ is not available for the agent at w. $\bigcup_{\delta \in Act} R_\delta(w)$ is the set of possible successor states of w, depending on how the agent and the environment are going to behave. Since $\bigcup_{\delta \in Act} R_\delta$ is not necessarily deterministic, we can model branching in **CGS**: the very same state may have different successors, depending on the behaviour of the agent and the environment. Since $\bigcup_{\delta \in Act} R_\delta(w) \neq \emptyset$ for any state $w \in W$, there should be at least one available action for the agent at any state.[17]

We model histories in **CGS** via *traces*.

Definition 8 (Traces) Given a **CGS** $M = (W, Act, \{R_\delta\}_{\delta \in Act}, V)$, a trace is a pair $\tau = (\tau_S, \tau_A)$, where $\tau_S : \mathbb{N} \to W$, $\tau_A : \mathbb{N} \to Act$, such that for every $n \in \mathbb{N}$: $\tau_S(n) R_{\tau_A(n)} \tau_S(n+1)$. In other words, a trace can be seen as an infinite path $w_1 R_{\delta_1} w_2 R_{\delta_2} \ldots$, where w_{n+1} is a possible outcome of doing δ_n at w_n for any $n \in \mathbb{N}$.

Given a trace $\tau = (\tau_S, \tau_A)$ and an integer $n \in \mathbb{N}$, by $\tau^{\geq n}$ we denote a trace $\kappa = (\kappa_S, \kappa_A)$, such that $\kappa_S(l) = \tau_S(n+l)$ and $\kappa_A(l) = \tau_A(n+l)$. I.e., $\tau^{\geq n}$ is a suffix of a path that starts with the $1+n$th state. For instance, $\tau^{\geq 0} = \tau$ for any trace τ. The set of all traces is denoted as Tr.

Given two traces $\tau = (\tau_S, \tau_A)$, $\kappa = (\kappa_S, \kappa_A)$, τ and κ are state-equivalent, $\tau \equiv_S \kappa$, iff $\tau_S(0) = \kappa_S(0)$; τ and κ are action-equivalent, $\tau \equiv_A \kappa$, iff $\tau \equiv_S \kappa$ and $\tau_A(0) = \kappa_A(0)$. For brevity, $[\tau]_S = \{\kappa \in Tr \mid k \equiv_S \tau\}$, $[\tau]_A = \{\kappa \in Tr \mid k \equiv_A \tau\}$ for any $\tau \in Tr$.

Informally, equivalence classes under $\equiv_S$ correspond to the moments in standard branching-time structures, while traces correspond to histories. Given any CGS, we can transform it into a discrete **BT**-structure via the unraveling technique, while every discrete **BT**-structure can already be seen as a CGS, where moments are states and histories are traces. Therefore, CGSs and discrete **BT**-structures are equivalent: for a detailed proof of this fact, see [31, Theorem 5.7].

Next, we add the partial set-selection function to model (conditional) intentions of the agent in the branching time environment. The function will enjoy all the good properties we have given to it in Sec. 4. In CGSs, the set of state-equivalent traces will replace the space of practically possible worlds: given any state w, a trace $\tau = (\tau_S, \tau_A)$ is practically possible for our agent (i.e., the future it represents may be realized given what actions are available for the agent at w) iff $\tau_S(0) = w$. As we are to see, truth of formulae are defined on traces instead of possible worlds, so propositions are set of traces instead of sets of possible worlds. Hence, the partial set-selection function gets a signature $f : (Tr \times 2^{Tr}) \to 2^{Tr}$. We also impose two additional properties. First, the *control constraint* holds: if the agent intends that φ conditional

[17] If we want to model situations, where the agent may lose their agency at some states (e.g., for a person to fall into a coma or for a computation to terminate), we can just add a designated action $\varepsilon \in Act$, to do nothing.

on ψ, then the agent should have an available action to bring about that φ conditional on ψ.[18] Second, *intentions are up to the agent*: if the agent intends something, then she has chosen that she intends so.

Definition 9 (Concurrent game structures with intentions) An intentional concurrent game structure is a tuple $M = (W, Act, \{R_\delta\}_{\delta \in Act}, f, V)$, where:

1. $(W, Act, \{R_\delta\}_{\delta \in Act}, V)$ is a **CGS** in accordance with Def. 7;
2. $f : Tr \times 2^{Tr} \to 2^{Tr}$ is a partial set-selection function with the following properties:
   (a) If $f(\tau, P)$ is defined, then $[\tau]_S \cap f(\tau, P) \cap P \neq \emptyset$: agent's intentions in case of P are consistent with P at the given moment;
   (b) If $f(\tau, P)$ is defined, then $f(\tau, P) \subseteq [\tau]_S$: the object of agent's present-directed conditional intention is a proposition about what the agent can bring about *now*, conditional on her and environment's behaviour;
   (c) If $f(\tau, P)$ and $f(\tau, Q)$ are defined, while $P \cap Q \cap [\tau]_S \neq \emptyset$, then $f(\tau, P \cap Q)$ is defined;
   (d) If $f(\tau, P)$ and $f(\tau, P \cap Q)$ are defined and $f(\tau, P) \cap P \cap Q \neq \emptyset$, then $f(\tau, P \cap Q) \subseteq f(\tau, P)$;
   (e) Given that $f(\tau, P)$ is defined, if $f(\tau, P) \subseteq Q$ and $f(\tau, P \cap Q)$ is defined, then $f(\tau, P) = f(\tau, P \cap Q)$;
   (f) *Control constraint*: if $f(\tau, P) = Q$, then there exists an action $\delta \in Act$, such that: there is a trace $\kappa \in P$, such that $\kappa \equiv_S \tau$ and $\kappa_A(0) = \delta$, and for any trace $\lambda \in [\kappa]_A$, if $\lambda \in P$, then $\lambda \in Q$;
   (g) *Intentions are up to the agent*: if $\tau \equiv_A \kappa$, then $f(\tau, P) = f(\kappa, P)$.

As we see, f function has all the properties as in Def. 1. Additionally, we require (2a) to treat state-equivalent traces as practically possible worlds and (2b) to make sure we model present-directed (conditional) intention. Control constraint expressed in (2f) requires that if the agent intends Q conditional on P, then the agent must be able to bring about Q conditional on P. To be more precise, if $f(w, P) = Q$, then there exists an action $\delta \in Act$, which is compatible with P and which guarantees that Q in case of P. We require this action to be compatible with the condition to avoid the situations, where the agent has a conditional intention, but can "realize" it only by preventing the condition from happening. Condition (2g) expresses that the agent performs her atomic actions with the same intention, independently of its outcome.

Definition 10 (Semantics for $\mathscr{L}_t$) Given a model $M = (W, Act, \{R_\delta\}_{\delta \in Act}, f, V)$ and a trace $\tau = (\tau_S, \tau_A) \in Tr$, the satisfiability relation $\models$ is inductively defined as follows:

$$M, \tau \models p \Leftrightarrow \tau_S(0) \in V(p)$$
$$M, \tau \models \neg\varphi \Leftrightarrow M, \tau \not\models \varphi$$
$$M, \tau \models \varphi \vee \psi \Leftrightarrow M, \tau \models \varphi \text{ or } M, \tau \models \psi$$
$$M, \tau \models \Box\varphi \Leftrightarrow \forall \kappa \in [\tau]_S : M, \kappa \models \varphi$$
$$M, \tau \models X\varphi \Leftrightarrow M, \tau^{\geq 1} \models \varphi$$

[18] The control constraint for unconditional intention is widely accepted among philosophers of action and defended in, among others, the following works: [10, 11, 17, 32]. It is natural to adopt the analogue constraint for conditional intentions, given their function as contingent commitments to actions.

$M, \tau \models \varphi U \psi \Leftrightarrow \exists n \in \mathbb{N} : M, \tau^{\geq n} \models \psi$ and $\forall m \in \mathbb{N}(0 \leq m \leq n \Rightarrow M, \tau^{\geq m} \models \varphi)$

$M, \tau \models [a]\varphi \Leftrightarrow \forall \kappa \in [\tau]_A : M, \kappa \models \varphi$

$M, \tau \models Int^{\varphi} X \psi \Leftrightarrow f(\tau, [\![\varphi]\!]_M)$ is defined and $f(\tau, [\![\varphi]\!]_M) \subseteq [\![X\psi]\!]_M$

where $[\![\varphi]\!]_M = \{\kappa \in Tr \,|\, M, \kappa \models \varphi\}$.

Example 2 (The rent contract) Anne is homeless and looking for a place to rent in Amsterdam. Since she struggles to find anything long-term, she considers *antikraak*.[19] Anne gets a proposition via antikraak to rent a place from a house owner who has temporarily left Amsterdam for Brussels. If Anne agrees, as soon as the house owner gets bored in Brussels and wants to come back to Amsterdam, Anne will have to leave the house.

Anne constantly gets information about the housing market from the insiders, so she can act conditionally on the updates she might receive. Anne intends to agree to rent via antikraak, in case no indefinite-term lease offers are coming on the market for the next month. Anne also intends to deny the antikraak proposal and wait for a lease for an indefinite term, if such an offer appears on the market in the next month.

Let $Var = \{m, h, p\}$, where m stands for *Anne has to **m**ove out*, h *– Anne has a **h**ouse to stay in*, p *– renting **p**roperty for an indefinite term appears on the market*.

We construct the following model for Example 2. Let $Act = \{w, a, \varepsilon\}$, where w is an action to reject antikraak and **w**ait for the indefinite term proposal; a –to **a**ccept antikraak, ε – to do nothing. We build a concurrent game model for the example as on Fig. 1.

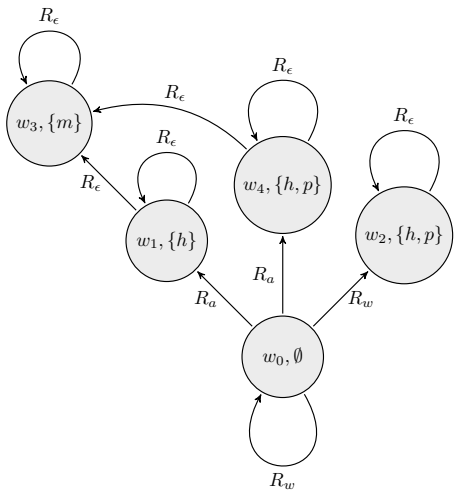

Fig. 1 Concurrent game model for Example 2.

Let w_0 be a world, where Anne receives the antikraak proposal and may either accept it or reject it and wait for the indefinite term proposal; w_1 – the world where Anne accepts

[19] Antikraak is a property-management practice in the Netherlands where people rent spaces in temporally vacant buildings to prevent squatting. The rent is typically cheap and the contract is easier to get, but the tenant has to move out on a short notice as soon as the owner needs to use the building again, so renting via antikraak presupposes unpredictability of terms.

antikraak, while no indefinite term proposals arrive the market; w_2 – the world where Anne rejected antikraak and kept waiting for the indefinite term proposal, and the proposal appears on the market; w_3 – the world were Anne has to move out, because the home owner got bored in Brussel; w_4 – the world where Anne accepts antikraak, while the indefinite term proposal arrives on the market.

Let τ be any trace such that $\tau_S(0) = w_0$. Let $P = [\![Xp]\!]_M$, and $Q = [\![X\neg p]\!]_M$. Then, $f(\tau,P) = \{\kappa \in Tr \mid \kappa \equiv_S \tau, \kappa_S(1) = w_2\}$; $f(\tau,Q) = \{\kappa \in Tr \mid \kappa \equiv_S \tau; \kappa_S(1) = w_1\}$. Then, Anne intends to rent the house indefinitely in case she gets an undetermined offer, and she intends to accept antikraak and stay there until she has to move out in case there is no indefinite term offer: $M,\tau \models Int^{Xp}X(Gh) \wedge Int^{X\neg p}X(hUXm)$.

Note that because of the control constraint, Anne cannot, e.g., unconditionally intend to rent a house for the next 6 months – $IntX(X^6h)$, where $X^1\varphi := X\varphi$, $X^{n+1}\varphi := X(X^n\varphi) \wedge X^n\varphi$ – since there is no action of Anne to guarantee that. Anne has two actions at w_0: a and w. If Anne chooses a at w_0, depending on how long the homeowner will stay in Brussels, she will or will not have a house for 6 months. If she chooses w at w_0, she might or might not get a house at all, depending on whether there will be an undetermined offer on the market. Moreover, Anne cannot intend to rent a house for the next 6 months even conditionally on her getting an undetermined offer or the homeowner staying in Brussels for 6 months. Let the condition be expressed as follows: $\varphi := X^7(\neg m) \vee Xp$, i.e. either the homeowner is not going back to Amsterdam for the next 7 months, or a permanent contract becomes available during this month. Still, $Int^{\varphi}X(X^6h)$ cannot be satisfied on **CGS** in Figure 1 with any well-defined function f. No single action of Anne guarantees $X(X^6h)$ in case of φ. If there will be no permanent contract but the homeowner is not going back to Amsterdam, then Anne should choose a; if a permanent contract becomes available but the homeowner is coming back to Amsterdam, Anne should choose w. Yet no action will guarantee $X(X^6h)$ in any case where φ holds.

Next, we list some interesting (in)validities on the class of CGSs with intentions. First, note that axioms and rules of **L**, adequately translated from $\mathscr{L}$ to $\mathscr{L}_i$, are still valid:

Proposition 5 *The following formulae are valid in the class of concurrent game structures with intentions:*

$$\models Int^{\varphi}X\psi \to [C]\varphi$$
$$\models Int^{\varphi}X(\psi \to \theta) \to (Int^{\varphi}X\psi \to Int^{\varphi}X\theta)$$
$$\models (\Box X\psi \wedge [C]\varphi) \to Int^{\varphi}X\psi$$
$$\models Int^{\varphi}X\psi \to \Diamond(\varphi \wedge X\psi)$$
$$\models ([C]X\alpha \wedge [C]X(\alpha \wedge \psi)) \to (Int^{X\alpha}X\psi \to (Int^{X\alpha}X\varphi \leftrightarrow Int^{X(\alpha\wedge\psi)}X\varphi))$$
$$\models ([C]X\alpha \wedge [C]X(\alpha \wedge \beta)) \to \big((Int^{X\alpha}X\varphi \wedge \neg Int^{X\alpha}X(\alpha \to \neg\beta)) \to Int^{X(\alpha\wedge\beta)}X\varphi\big)$$
$$\models \big([C]\varphi \wedge [C]\psi \wedge \Diamond(\varphi \wedge \psi)\big) \to [C](\varphi \wedge \psi)$$

Proof Follows from (2a-2e) of Def. 9, analogously to the proof of **L**'s soundness in Theorem 4. □

The following validities concern the interaction between agents' intentions, actions, and abilities:

Proposition 6 *The following formulae are valid in the class of concurrent game structures with intentions:*

$$\models Int^{\varphi} X\psi \leftrightarrow [a] Int^{\varphi} X\psi \tag{13}$$

$$\models Int^{\varphi} X\psi \rightarrow \Diamond[a]\big(\langle a\rangle\varphi \wedge (\varphi \rightarrow X\psi)\big) \tag{14}$$

Proof (13) corresponds to (2g), and (14) – to (2f) of Def. 9. □

Informally, (14) expresses the control constraint: if the agent intends that $X\psi$ conditional on φ, then she has an action to guarantee that $X\psi$ if φ that is compatible with φ. (13) expresses the fact that intentions are up to the agent: she intends something iff she sees to it that she so intends.

In our formalism, we do not presuppose the persistence of intentions, since they are understood as present-directed. For instance, in [11], it is argued that if the idealized rational agent intends that φ, then she must continue to intend so until she reaches the state where either φ is true or no longer possible for her to achieve. In $\mathscr{L}_t$, this can be expressed by the following formula: $IntX(F\varphi) \rightarrow (IntX(F\varphi))U(\varphi \vee \Box G\neg\varphi)$. However, this formula is not valid in the class of CGSs with intentions. Given that we interpret $IntX(F\varphi)$ as a present-directed intention, this is intuitively plausible. For example, assume I have a present-directed intention that my colleague will eventually die from poisoning. Then, according to (14), I should have an available action that would immediately satisfy my intention, e.g., I can poison her coffee. After I poisoned her coffee, I would have known that she would eventually die from poisoning, so there would be no need for me to keep my intention until she actually died. I would have achieved what I aimed for, but I would not have seen the desired consequences yet. If I, despite my intention, failed to poison my colleague's coffee, I still might drop that intention: it could be that I had reasons to poison my colleague only at that past moment, but in the present, there is no reason for me to do so. On the contrary, if I have a future-directed intention that my colleague will eventually die from poisoning, then my intention should persist until I either reach my goal or realize that the goal is no longer possible to reach.

We proceed with showing which important notions and distinctions can be expressed in $\mathscr{L}_t$ as opposed to $\mathscr{L}$. In $\mathscr{L}_t$, we can track whether the agent has control over the conditions of her intentions. This is important, since, as Ludwig argues, whenever the agent has control over the condition and exercises it, the corresponding conditional intention should be reducible to the unconditional one:

Proposition 7 (Control over conditions, Ludwig's thesis) *"If conditions are in an agent's control, they can be antecedents only if the agent has sufficient reason not to exercise control (negative or positive as the case may be) over their obtaining" [1, p.43].*

The positive control over φ means the ability to ensure that φ holds: $\Diamond[a]\varphi$; the negative control over φ means the ability to prevent φ: $\Diamond[a]\neg\varphi$.

Often it makes sense to have an intention conditional on φ, while having only negative or only positive control over φ and refusing to exercise it. For instance, a nurse has only positive control over her patient's death and only negative control over his survival (i.e., she can kill him, but cannot ensure that he will survive). She may, e.g., intend to change his dressing in

case he survives and intend to notify the doctor in case he dies. In both cases, whether the patient dies or survives is a contingency for her, given that she does not intend to kill the patient. If the nurse has both positive and negative control over the patient's life, it seems rational for her to have such conditional intentions if she refuses to exercise her control over the patient's life, i.e., if she intends neither to kill nor to save him. If there are no reasons not to decide the patient's fate, intentions conditional on the decision she intends to make are reducible to her unconditional intentions via the (Cut) axiom.

Notably, we can distinguish *pre*conditional intentions in $\mathscr{L}_t$. $Int^{\varphi} X\psi \wedge \Diamond\neg[a]X\psi$ stands for "*the agent intends that* ψ *conditional on* φ *and is unable to ensure* ψ *unconditionally*". As we have noted before, $Int^{\varphi} X\psi$ entails $\Diamond[a]\big(\langle a\rangle\varphi \wedge (\varphi \to X\psi)\big)$. Hence, the agent intends that ψ conditional on φ, can ensure ψ conditionally on φ, but cannot ensure it unconditionally. So φ entails a necessary precondition for that agent to ensure ψ. Moreover, sometimes we can separate reasons from preconditions when they are mixed. For instance, assume one intends to go to a movie theater, in case a good film is screened and she has enough money to buy a ticket. Then, the condition contains a reason (the good film is screened) and a precondition (one has enough money to get a ticket) at the same time. In $\mathscr{L}_t$, we can express the difference. Let φ, ϑ and ψ stand for "*the good film is screened*", "*one has enough money to get a ticket*" and "*one watches the film in the movie theater*", respectively. Then, the intention in question is expressible as follows: $Int^{(\varphi\wedge\vartheta)} X\psi$. One can watch a movie in the movie theater in case she has enough money, regardless of whether the film is good or not:

$$\Diamond[a]\big(\langle a\rangle\vartheta \wedge (\vartheta \to X\psi)\big)$$

Yet she cannot watch a film unconditionally, since she needs money to buy a ticket: $\neg\Diamond[a]X\psi$. Hence, in the condition $(\varphi\wedge\vartheta)$, φ is a reason, while ϑ is a precondition.

To conclude this subsection, we have provided a richer formalism with the explicit treatment of action and time, where we understand conditional intention as present-directed. In this formalism, we can clearly see with which conditional intention the agent acts, which is useful for formal models of *mens rea*. We have modeled intentions as present-directed, parting ways with the seminal Cohen and Levesque's work, to adjust our formalism for exactly that use-case. We can potentially develop a similar formalism for future-directed intentions: we would need to drop the constraint (2b); in the formulation of control constraint ((2f) of Def. 9), we would need to replace the existence of an atomic action with the existence of a long-term *strategy*, as well as add new constraints to account for persistence of conditional intentions. We leave this investigation for the future work.

5.2 Side-effects problem

The logic **L** suffers from the problem of *logical omniscience*: conditional intention is closed under necessary entailment.

$$Int^{\varphi}\psi \wedge \Box(\psi \to \vartheta) \models Int^{\varphi}\vartheta \tag{15}$$

Hence, our formalism cannot handle the problem of *side-effects*. By a side effect we mean an outcome of the intended action, which the agent does not intend to achieve. As philosophical [17, 33], psychological and experimental [34, 35] investigations show, agents do not intend all the consequences of their intended actions and are not rationally pressured to do so, while the

difference between intended and unintended consequences of intentional actions is crucial, especially in normative reasoning.

In general, an agent can fail to intend some a priori consequence of her intention for two reasons [36]:

1. She lacks cognitive resources to grasp the consequence or compute the proof that the consequence indeed follows from what she intends. For instance, one may intend to write down Peano axioms without intending to write down an axiomatic system in which the theorem of prime numbers is provable, given that one does not know how to derive the theorem from Peano axioms.
2. The consequence in question is irrelevant to the agent's practical matters. For instance, while on the trip to France, one might intend to buy a bottle of wine in a legally working store to avoid getting uncertified products. Buying goods from a legally operating store in France practically entails paying some VAT in the budget of France. Even knowing that entailment, one still might not intend to pay some money to the French budget, simply because it is completely irrelevant to her practical matters.

We can abstract away from the first kind of failures of the closure under necessary entailment by assuming that the agent whose intentions we model is a perfect logician: she is unbounded in her cognitive capacities when it comes to logical reasoning. The second kind of closure's failures is more interesting. We can demonstrate how our logic suffers from the inability to model them by reformulating the famous gentle murder paradox [37] in terms of intention. The following statements seem to be consistent with one another:

I intend not to kill anyone.	(S1)
It is not true that I intend to kill in case I am killing.	(S2)
In case I am killing someone, I intend to do it gently.	(S3)
If I kill someone gently, it necessarily follows that I kill.	(S4)

If we formalize (S1)-(S4) in our logic, we will get the next set of formulae:

$$\{Int\neg kill, \neg Int^{kill} kill, Int^{kill} gentle, \Box(gentle \to kill)\}$$

By (15), $Int^{kill} gentle, \Box(gentle \to kill) \models Int^{kill} kill$, hence,

$$\{Int\neg kill, \neg Int^{kill} kill, Int^{kill} gentle, \Box(gentle \to kill)\} \models \neg Int^{kill} kill \wedge Int^{kill} kill$$

Therefore, in our logic we cannot handle the gentle murder paradox: we cannot model situations where one is unintentionally committing a murder, while intentionally doing it as gently as possible.

Since the problem of side-effects is not the primary aim of this paper, we have abstracted away from it to keep our framework simple. We can alleviate the problem by either implementing some raw syntactical restrictions of the closure, using machinery of awareness functions [38] as it was done in [8, 39], or taking a more subtle solution, using a question-sensitive framework [10, 36]. To keep things simple, we show a toy solution in the spirit of

[38]. Namely, we do not explicitly model why some consequences of one's intentions are unintended and simply restrict the closure via the analogue of awareness functions.

First, we expand our language with *explicit intention* modality:

$$\mathscr{L}_I \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \vee \varphi) \mid \Box\varphi \mid Int^{\varphi}\psi \mid \mathscr{I}^{\varphi}\psi$$

where $Int^{\varphi}\psi$ reads as "ψ *is entailed by what the agent intends in case of* φ" (from now on, will be abbreviated to "*agent implicitly intends that* ψ *in case of* φ"), $\mathscr{I}^{\varphi}\psi$ – "*agent explicitly intends that* ψ *in case of* φ". Explicit unconditional intentions are expressible via conditional ones standardly: $\mathscr{I}\varphi := \mathscr{I}^{\top}\varphi$.

As for the semantics, we will extend conditional frames with *availability functions*.

Definition 11 (Hyperintensional conditional models) A class of hyperintensional conditional frames **H** consists of tuples of the form $F = (W, f, a)$, where:

1. (W, f) is a conditional frame as in Definition 1;
2. $a : W \times 2^W \to 2^{\mathscr{L}_I}$ is a partial availability function[20] that takes a world w, a proposition P and maps it to the set of formulae that are available for the agent in w to intend in case of P. a has the following properties:
   (a) Definedness: $a(w,P)$ is defined iff $f(w,P)$ is defined;
   (b) Restricted monotonicity: given that $a(w,P)$ and $a(w,P\cap Q)$ are defined, if $f(w,P)\cap P\cap Q \neq \emptyset$, then $a(w,P) \subseteq a(w,P\cap Q)$;
   (c) Cut: given that $a(w,P)$ and $a(w,P\cap Q)$ are defined, if $f(w,P) \subseteq Q$, $a(w,P) = a(w,P\cap Q)$.

As usual, any frame $F = (W, f, a)$ can be extended to a model $M = (F, V)$, where V is a standard valuation function.

Availability function a specifies what formulae the agent can conditionally intend. It takes a world and a proposition $P \subseteq W$, since the very same formulae may or may not be available, depending on the condition of the intention. For example, it is irrational to unconditionally intend to win the lottery, since it is out of the player's control, but one can intend to win the lottery in case one gets to know which lottery ticket is the winning one. Property 2a means that one can intend something conditional on the proposition $P \subseteq W$ iff P is a condition for one's intention. Properties 2b and 2c ensure that the monotonicity principles we have argued for in the previous sections are valid for explicit intentions.

Definition 12 (Semantic clauses for $\mathscr{L}_I$) For any hyperintensional conditional model $M = (W, f, a, V)$ and any world $w \in W$, $\models$ relation inductively defined as follows (all cases, except of $\mathscr{I}^{\varphi}\psi$, are identical to Def. 2):

$$M, w \models \mathscr{I}^{\varphi}\psi \Leftrightarrow M, w \models Int^{\varphi}\psi \text{ and } \psi \in a(w, [\![\varphi]\!])$$

[20] In [38], such functions are called *awareness* functions, and $\varphi \in a(w)$ is supposed to indicate that the agent is aware of φ in w. Here, we purposefully avoid this name, since an agent being aware of the proposition is not enough for that proposition to be available for the agent to intend. See discussions around cases of foreseeing without intending, where agents foresee (and hence are aware of) some consequences of their intentions, but still do not intend them [10, 17, 40].

The notions of $M \models \varphi$, $F, w \models \varphi$, $F \models \varphi$ and $\mathbf{H} \models \varphi$ are defined standardly.

Using hyperintensional models, we can easily avoid the closure under entailment and hence alleviate the side-effects problem. Going back to the gentle murderer paradox, it is simple to handle. Take a model $M = (W, f, a, V)$, where:

- $W = \{w_1, w_2, w_3\}$, where in w_1 one is not killing anyone, in w_2 one is murdering someone gently and in w_3 one is murdering someone cruelly. Consequently, $V(kill) = \{w_2, w_3\}$, $V(gentle) = \{w_2\}$. Let the real world be w_1.
- $f(w_1, W) = \{w_1\}$, i.e. one intends not to kill anyone no matter what.
- $f(w_1, \{w_2, w_3\}) = \{w_2\}$, i.e. in case one is killing someone, one intends to do it gently.
- $a(w_1, W) = \{\neg kill, kill\}$, $a(w_1, \{w_2, w_3\}) = \{gentle, \neg gentle\}$ i.e. the agent can unconditionally intend either to kill or not to kill, but it case he kills, he can only intend either to do it gently or not: given that $f(w_1, W) \subseteq W \setminus \{w_2, w_3\}$, the intention in case he kills is a plan-B intention, hence, the agent finds himself in a situation of unintentional killing, and committing a murder is not his choice and cannot be intended.

Clearly, $M, w_1 \models \mathscr{I}\, \neg kill \wedge \mathscr{I}^{kill} gentle \wedge \neg \mathscr{I}^{kill} kill$, despite the fact that $M, w_1 \models \Box(gentle \to kill)$, since $kill \notin a(w_1, [\![kill]\!])$. Interestingly, if we changed the story and imagined that the agent in question unconditionally intends to kill, then he would intend to kill in case he kills as well: $\mathscr{I}\, kill, [C]kill \models \mathscr{I}^{kill} kill$. This would happen due to the properties of availability functions we have established. This is desirable, since then the plan for the case he commits a murder is not a plan-B for him, but the very same plan as the unconditional one, hence, he should explicitly intend the same things.

Since we are abstracting away from the exact reasons why some formulae cannot be explicitly intended by the agent, we do not impose any further conditions on availability functions. Here we list what conditions should be added to satisfy apparently plausible closure principles. The closure principles are on the right, while the conditions of availability functions are on the left.[21] In what follows, $\circ \in \{\wedge, \vee\}$:

$\psi_1 \circ \psi_2 \in a(w, [\![\varphi]\!]) \Leftrightarrow \psi_2 \circ \psi_1 \in a(w, [\![\varphi]\!])$	iff $\mathscr{I}^{\varphi}(\psi_1 \circ \psi_2) \leftrightarrow \mathscr{I}^{\varphi}(\psi_2 \circ \psi_1)$
$\psi_1 \wedge \psi_2 \in a(w, [\![\varphi]\!]) \Leftrightarrow \psi_1, \psi_2 \in a(w, [\![\varphi]\!])$	iff $\mathscr{I}^{\varphi}(\psi_1 \wedge \psi_2) \leftrightarrow (\mathscr{I}^{\varphi}\psi_1 \wedge \mathscr{I}^{\varphi}\psi_2)$
$\psi_1, \psi_2 \in a(w, [\![\varphi]\!]) \Rightarrow \psi_1 \vee \psi_2 \in a(w, [\![\varphi]\!])$	iff $(\mathscr{I}^{\varphi}\psi_1 \wedge \mathscr{I}^{\varphi}\psi_2) \to \mathscr{I}^{\varphi}(\psi_1 \vee \psi_2)$
$\neg\neg\psi \in a(w, [\![\varphi]\!]) \Leftrightarrow \psi \in a(w, [\![\varphi]\!])$	iff $\mathscr{I}^{\varphi}\neg\neg\psi \leftrightarrow \mathscr{I}^{\varphi}\psi$
If $\models \psi$, then $\psi \notin a(w, [\![\varphi]\!])$	iff If $\vdash \psi$, then $\vdash \neg\mathscr{I}^{\varphi}\psi$

Using availability functions, we can also adjust our logic for an opponent of propositionalism about intention. Namely, assume we have a set of propositions $Var = \{p_1, p_2, \ldots\}$ with a distinguished subset of *agentives* $Ag \subseteq Var$, i.e. propositions about the agent's actions. Restricting the range of a function to modal-free formulae built from agentives, we can avoid

[21] We emphasize the meaning of the last condition on the availability functions. Namely, it prevents the agent from explicitly intending logical truths. This condition was advocated, among others, by Cohen and Levesque, who viewed intentions as a certain kind of *achievement goals* that are believed to be currently false [11, Def.3.24]. A rational agent should not perceive logical truths as false and aim at making them true. We are grateful to the anonymous reviewer for pointing out this important condition.

situations where the object of the agent's intention is a proposition about anything but her own actions.

If we do not impose any of those additional properties on a, the logic $\mathbf{L}_I = \{\varphi \in \mathscr{L}_I \mid \mathbf{H} \models \varphi\}$ of hyperintensional conditional frames $\mathbf{H}$ can be axiomatized as follows: see Table 2.

($\mathbf{L}$)	Axioms and rules of $\mathbf{L}$
(Inc)	$\mathscr{I}^{\varphi}\psi \to Int^{\varphi}\psi$
(Ex-Cut)	$([C]\varphi \wedge [C](\varphi \wedge \psi)) \to (Int^{\varphi}\psi \to (\mathscr{I}^{\varphi}\theta \leftrightarrow \mathscr{I}^{(\varphi\wedge\psi)}\theta))$
(R-ExMon)	$([C]\alpha \wedge [C](\alpha \wedge \beta)) \to (\neg Int^{\alpha}(\alpha \to \neg\beta) \to (\mathscr{I}^{\alpha}\varphi \to \mathscr{I}^{(\alpha\wedge\beta)}\varphi))$
(RE)	If $\vdash \varphi \leftrightarrow \psi$, then $\vdash \mathscr{I}^{\varphi}\theta \leftrightarrow \mathscr{I}^{\psi}\theta$

Table 2 Logic $\mathbf{L}_I$

The proof of the fact that $\mathbf{L}_I$ is sound and complete w.r.t. $\mathbf{H}$ strongly resembles the proof of Theorem 4 with the slight modification of canonical model construction. From that point, by following the steps of completeness proof of $\mathbf{L}$ w.r.t. $\mathbf{C}$, one may get the analogous result for $\mathbf{L}_I$ and $\mathbf{H}$.

We still want to construct a finite canonical model for every formula ϕ, but we cannot use the same strategy with restricted languages anymore, since $\mathscr{I}$ modality does not validate the rule of equivalence. From $\vdash \varphi \leftrightarrow \psi$ it does not follow that $\vdash \mathscr{I}^{\theta}\varphi \leftrightarrow \mathscr{I}^{\theta}\psi$, therefore, if we construct the restricted language for ϕ the same way we did it in Section 4, we will have infinite mcs in that language. Hence, we will redefine the restricted language to avoid it.

Definition 13 (Restricted sub-languages of $\mathscr{L}_I$) Given any formula $\phi \in \mathscr{L}_I$, let $sub(\phi)$ denote the set of subformulae of ϕ. Then, the set of formulae Σ_I^{ϕ} is defined as follows:

$$\Sigma_I^{\phi} \ni \psi ::= \pi \mid \neg\psi \mid (\psi \vee \psi) \mid \Box\psi \mid Int^{\psi}\psi \mid \mathscr{I}^{\psi}\pi$$

where $\pi \in sub(\phi)$. Note that for any $\psi \in \Sigma_I^{\phi}$, $Var(\psi) \subseteq Var(\phi)$ and for any $\psi \in \Sigma_I^{\phi}$ there are finitely many $\mathscr{I}^{\psi}$-formulae. Then, we define a restricted language $\mathscr{L}_I^{\phi}$ as follows:

$$\mathscr{L}_I^{\phi} = \{\psi \in \Sigma_I^{\phi} \mid md(\psi) \leq md(\phi)\}$$

where md function is defined analogously to Def. 3, i.e. all the induction steps for Boolean connectives, $\Box$ and Int are the same and

$$md(\mathscr{I}^{\varphi_1}\varphi_2) = 1 + max(md(\varphi_1), md(\varphi_2))$$

Definition 14 (Canonical model for $\mathbf{L}_I$) Let $MCS_{\phi} \subseteq 2^{\mathscr{L}_I^{\phi}}$ be a collection of mcs of the restricted language $\mathscr{L}_I^{\phi}$. Then, given any $\Gamma \subseteq MCS$, a canonical model is a tuple $M_{\Gamma}^c = (W_{\Gamma}^c, f_{\Gamma}^c, a_{\Gamma}^c, V_{\Gamma}^c)$, where:

- $(W_{\Gamma}^c, f_{\Gamma}^c, V_{\Gamma}^c)$ are defined just like in Def. 5;
- $a_{\Gamma}^c(\Delta, P) = \begin{cases} \{\psi \in \mathscr{L} \mid \mathscr{I}^{\varphi}\psi \in \Delta\} & \text{if } P = \widehat{\varphi}, [C]\varphi \in \Delta \\ \text{undefined} & \text{otherwise} \end{cases}$

Lemma 4 (Truth lemma for M^c) Given any restricted language $\mathscr{L}_I^\phi$ and a canonical model $M_\Gamma^c = (W_\Gamma^c, f_\Gamma^c, a_\Gamma^c, V_\Gamma^c)$, where $\Gamma \in MCS_\phi$, any mcs $\Delta \in W_\Gamma$ and any formula $\varphi \in \mathscr{L}_I^\phi$, $M_\Gamma^c, \Delta \models \varphi$ iff $\varphi \in \Delta$.

Proof By induction on the complexity of φ. All cases, except of $\mathscr{I}$-formulae, were proven in Lemma 3. Assume that $\varphi := \mathscr{I}^{\psi_1}\psi_2$. Left-to-right. $M_\Gamma^c, \Delta \models \mathscr{I}^{\psi_1}\psi_2$. I.e. $f_\Gamma^c(\Delta, [\![\psi_1]\!]) \subseteq [\![\psi_2]\!]$ and $\psi_2 \in a_\Gamma^c(\Delta, [\![\psi_1]\!])$. By induction hypothesis, $f_\Gamma^c(\Delta, \widehat{\psi_1}) \subseteq \widehat{\psi_2}$ and $\psi_2 \in a_\Gamma^c(\Delta, \widehat{\psi_1})$. Then, there exists some formula θ, such that $\widehat{\psi_1} = \widehat{\theta}$ (i.e. $\Box\Gamma \vdash \psi_1 \leftrightarrow \theta$), and $[C]\theta, \mathscr{I}^\theta\psi_2 \in \Delta$. By (RE) it follows that $\mathscr{I}^{\psi_1}\psi_2 \in \Delta$. Right-to-left: Assume that $\mathscr{I}^{\psi_1}\psi_2 \in \Delta$. Then, by definition of a_Γ^c, $\psi_2 \in a_\Gamma^c(\Delta, \widehat{\psi_1})$, i.e., by induction hypothesis, $\psi_2 \in a_\Gamma^c(\Delta, [\![\psi_1]\!])$. Since $\mathscr{I}^{\psi_1}\psi_2 \in \Delta$, by (Inc), $Int^{\psi_1}\psi_2 \in \Delta$, from which by Lemma 3 it follows that $f_\Gamma^c(\Delta, [\![\psi_1]\!]) \subseteq [\![\psi_2]\!]$. Taking it all together, $M_\Gamma^c, \Delta \models \mathscr{I}^{\psi_1}\psi_2$. □

Lemma 5 Take any formula $\phi \in \mathscr{L}_I$. Let $M_\Gamma^c = (W_\Gamma^c, f_\Gamma^c, a_\Gamma^c, V_\Gamma^c)$ be a canonical model for some $\Gamma \in MCS_{\mathscr{L}_I^\phi}$. Then, M_Γ^c is a **H**-model.

Proof From Lemma 1 we know that $(W_\Gamma^c, f_\Gamma^c, V_\Gamma^c)$ is a **C**-model. Hence, it is left for us to prove that $a_\Gamma^c : W_\Gamma \times 2^{W_\Gamma} \to 2^{\mathscr{L}_I^\phi}$ is a partial availability function with the listed properties in Def. 11. It is a routine to check that a_Γ is a well-defined partial function of the corresponding type. As for properties:

1. $a_\Gamma^c(\Delta, P)$ is defined iff $f_\Gamma^c(\Delta, P)$ is defined. Follows directly from definitions of f_Γ^c and a_Γ^c.
2. Given that $a_\Gamma^c(\Delta, P)$ and $a_\Gamma^c(\Delta, P \cap Q)$ are defined, if $f_\Gamma^c(\Delta, P) \cap P \cap Q \neq \emptyset$, then $a_\Gamma^c(\Delta, P) \subseteq a_\Gamma^c(\Delta, P \cap Q)$. Assume that $a_\Gamma^c(\Delta, P)$ and $a_\Gamma^c(\Delta, P \cap Q)$ are defined and $f_\Gamma^c(\Delta, P) \cap P \cap Q \neq \emptyset$. Then, by definition of a_Γ^c, $P = \widehat{\varphi}$, $P \cap Q = \widehat{\varphi \wedge \psi}$ for some formulae $\varphi, \psi \in \mathscr{L}_I^\varphi$, such that $[C]\phi \in \Delta$ and $[C](\phi \wedge \psi) \in \Delta$. Then, $f_\Gamma^c(\Delta, P) \cap P \cap Q \neq \emptyset$ means that $f_\Gamma^c(\Delta, [\![\varphi]\!]) \cap [\![\varphi \wedge \psi]\!] \neq \emptyset$, from which it follows that $M_\Gamma^c, \Delta \models [C]\varphi \wedge [C](\varphi \wedge \psi) \wedge \neg Int^\varphi(\varphi \to \neg(\varphi \wedge \neg\psi))$. From Lemma 4 it follows that $[C]\varphi \wedge [C](\varphi \wedge \psi) \wedge \neg Int^\varphi(\varphi \to \neg(\varphi \wedge \neg\psi)) \in \Delta$. Then, by (R-ExMon) axiom, $\mathscr{I}^\varphi\vartheta \to \mathscr{I}^{(\varphi\wedge\psi)}\vartheta \in \Delta$ for any ϑ, from which it directly follows that $a_\Gamma^c(\Delta, \widehat{\varphi}) \subseteq a_\Gamma^c(\Delta, \widehat{\varphi \wedge \psi})$.
3. Given that $a_\Gamma^c(\Delta, \widehat{\varphi})$ and $a_\Gamma^c(\Delta, \widehat{\varphi \wedge \psi})$ are defined, if $f_\Gamma^c(\Delta, \widehat{\varphi}) \subseteq \widehat{\psi}$, then $a_\Gamma^c(\Delta, \widehat{\varphi}) = a_\Gamma^c(\Delta, \widehat{\varphi})$. Assume that $a_\Gamma^c(\Delta, \widehat{\varphi})$ and $a_\Gamma^c(\Delta, \widehat{\varphi \wedge \psi})$ are defined and $f_\Gamma^c(\Delta, \widehat{\varphi}) \subseteq \widehat{\psi}$. Then, by Definition 14, $M_\Gamma^c, \Delta \models [C]\varphi \wedge [C](\varphi \wedge \psi) \wedge Int^\varphi\psi$. From Lemma 3 it follows that $[C]\varphi \wedge [C](\varphi \wedge \psi) \wedge Int^\varphi\psi \in \Delta$. Then, by (Ex-Cut) axiom, $\mathscr{I}^\varphi\theta \leftrightarrow \mathscr{I}^{(\varphi\wedge\psi)}\theta$ for any $\theta \in \mathscr{L}_I^\phi$, from which it directly follows that $a_\Gamma^c(\Delta, \widehat{\varphi}) = a_\Gamma^c(\Delta, \widehat{\varphi \wedge \psi})$.

□

Theorem 8 *Given any $\varphi \in \mathscr{L}_I$, the following holds:*

$$\mathbf{H} \models \varphi \Leftrightarrow \mathbf{L}_I \vdash \varphi$$

Proof ($\Leftarrow$) By a routine check of the validity of axioms and the validity preservation of rules of inference.

($\Rightarrow$) By contraposition. Assume that $\mathbf{L}_I \nvdash \varphi$. Then, $\{\neg\varphi\}$ is $\mathbf{L}_I$-consistent. Take a language $\mathscr{L}_I^{\neg\varphi}$ and take any mcs $\Gamma \ni \neg\varphi$. Construct a canonical model $M_\Gamma^c = (W_\Gamma^c, f_\Gamma^c, a_\Gamma^c, V_\Gamma^c)$. Then, by Lemma 4, $M_\Gamma^c, \Gamma \models \neg\varphi$, while M_Γ^c is a **H**-model, as we know from Lemma 5. Hence, $\mathbf{H} \not\models \varphi$. □

5.2.1 Discussion: is (Ex-Cut) too strong?

One may argue that the axiom (Ex-Cut) is counterintuitive. Consider the next example:[22]

Example 3 Inne implicitly intends to eat lunch in case she finishes cooking. Inne explicitly intends to wash the dishes if she finishes cooking and eats lunch. But intuitively, Inne may not explicitly intend to wash the dishes afterwards if she finishes cooking: if she only knows that she finishes cooking, but it is undecided whether she will use the dishes while eating or not, it makes little sense to intend to wash them.

To put the example in line with our understanding of intentions as present-directed, we assume that Inne has the following four available actions at the moment of deliberation: not to cook her lunch (σ_1); to cook her lunch and not to eat it (σ_2); to cook her lunch, to eat it and not to wash the dishes afterwards (σ_3); and to cook her lunch, to eat it and to wash the dishes afterwards (σ_4). We abstract away from the compositional nature of these actions and treat them as atomic, so that the intentions in the example are indeed present-directed.

Let c, l, d be propositions that read as "*Inne finishes* ***c****ooking*", "*Inne eats* ***l****unch*", "*Inne washes the* ***d****ishes*", respectively. Then, the example presents an intuition for the invalidity of the following instance of (Ex-Cut):

$$[C]c \wedge [C](c \wedge l) \wedge Int^c l \wedge \mathscr{I}^{(c \wedge l)} d \not\rightsquigarrow \mathscr{I}^c d$$

In defense of the (Ex-Cut) axiom, we argue that Inne is rationally pressured to intend to wash dishes in case she is finished cooking, at least under the idealizations we have made. Apparently, Inne should not intend to wash her dishes if she cooks, if we take the intention out of context. It seems like Inne is uncertain whether she will eat and use her dishes or not in case she just cooks, so she should only intend to wash her dishes if she cooks and eats. But Inne implicitly intends to eat in case she cooks. Since we idealize Inne as practically omniscient, she knows that it follows from what she intends; hence, she is certain that she will eat in case she cooks. Consequently, whatever she intends in case she cooks and eats should be intended in case she just cooks, and vice versa.

If Inne might not know all the practical consequences of her intentions, she might not take it to be settled whether she eats in case she cooks or not, given that her intention is *implicit*. So if we drop the idealization about agents' practical omniscience, (Ex-Cut) is indeed too strong. We do commit to this idealization and endorse the (Ex-Cut) axiom, simply because bounded theoretical rationality is not in the scope of our interest. By introducing availability functions, our main goal is to rule out the side-effects that are practically irrelevant to the agent rather than just unexpected, given her inability to foresee some a priori consequences of her actions.

6 Conclusion

In this paper, we have introduced a logic of conditional intention. Our formalism inherits Bratmanian theory as a conceptual background, together with Ludwig and Ferrero's work on conditional intention. We have shown a number of properties of conditional intention that

[22] We are grateful to the anonymous reviewer for the counterexample.

are unique in the realm of conditional logics. Namely, agents are not rationally pressured to conditionalize their intentions on any proposition, conditional intentions should be restrictedly monotonic, and (under some reasonable limitations) free choice inferences should be valid. We have developed the semantical framework – conditional models **C** – that meets all these goals. We have also defined a logic **L**, which we proved to be sound and complete with respect to **C** and enjoys the finite model property. Moreover, we have discussed the ways to refine our framework so that we can handle connections with actions in time, as well as the side-effects problem.

We see a number of ways to continue this research. Namely, it would be interesting to look into intention revision. Conditional intentions may be seen as guides on how to revise one's intention in case some new information is learned. Explicit representation of this revision process and its comparison with similar frameworks, e.g., AGM-models or dynamics of unconditional intention [9] is a promising path to take. Looking into the multi-agent version of our logic is also interesting, since agents may conditionalize their intentions of actions or intentions on one another, and such interdependent intentions play an important role in coordination and team reasoning [41].

References

[1] Ludwig, K.: What are conditional intentions? Methode: Analytic Perspectives (4(6)), 30–60 (2015)

[2] Ferrero, L.: Conditional intentions. Noûs **43**(4), 700–741 (2009)

[3] Campbell, K.: Conditional intention. Legal Studies **2**(1), 77–97 (1982)

[4] Yaffe, G.: Conditional intent and mens rea. Legal Theory **10**(4), 273–310 (2004)

[5] Klass, G.: A conditional intent to perform. Legal Theory **15**(2), 107–147 (2009)

[6] Cohen, P.R., Levesque, H.J.: Persistence, intention, and commitment (1990)

[7] Konolige, K., Pollack, M.E.: A representationalist theory of intention. In: IJCAI, vol. 93, pp. 390–395 (1993)

[8] Rao, A.S., Georgeff, M.P.: Modeling rational agents within a bdi-architecture. Readings in agents, 317–328 (1997)

[9] Der Hoek, W., Jamroga, W., Wooldridge, M.: Towards a theory of intention revision. Synthese **155**(2), 265–290 (2007)

[10] Beddor, B., Goldstein, S.: A question-sensitive theory of intention. The Philosophical Quarterly **73**(2), 346–378 (2023)

[11] Cohen, P.R., Levesque, H.J.: Intention is choice with commitment. Artificial intelligence **42**(2-3), 213–261 (1990)

[12] Lorini, E., Herzig, A.: A logic of intention and attempt. Synthese **163**(1), 45–77 (2008)

[13] Sayre, F.B.: Mens rea. Harvard Law Review **45**(6), 974–1026 (1932)

[14] Broersen, J.: Deontic epistemic stit logic distinguishing modes of mens rea. Journal of Applied Logic **9**(2), 137–152 (2011)

[15] Lorini, E., Longin, D., Mayor, E.: A logical analysis of responsibility attribution: emotions, individuals and collectives. Journal of Logic and Computation **24**(6), 1313–1339 (2014)

[16] Abarca, A.I.R., Broersen, J.M.: A stit logic of responsibility. In: AAMAS, pp. 1717–1719 (2022)

[17] Bratman, M.: Intention, Plans, and Practical Reason. Harvard University Press (1987)

[18] Ferrero, L.: Ludwig on conditional intentions. Methode: Analytic Perspectives (6) (2015)

[19] Sverdlik, S.: Consistency among intentions and the 'simple view'. Canadian Journal of Philosophy **26**(4), 515–522 (1996)

[20] Zhu, J.: On the principle of intention agglomeration. Synthese **175**, 89–99 (2010)

[21] Chellas, B.F.: Basic conditional logic. Journal of philosophical logic, 133–153 (1975)

[22] Baltag, A., Smets, S.: Conditional doxastic models: A qualitative approach to dynamic belief revision. Electronic notes in theoretical computer science **165**, 5–21 (2006)

[23] Prakken, H., Sergot, M.: Contrary-to-duty obligations. Studia Logica **57**, 91–115 (1996)

[24] Van Der Torre, L.W.: Violated obligations in a defeasible deontic logic. In: ECAI, vol. 94, pp. 371–375 (1994). Citeseer

[25] Horty, J.F.: Moral dilemmas and nonmonotonic logic. Journal of philosophical logic **23**, 35–65 (1994)

[26] Owens, D.: Promises and conflicting obligations. J. Ethics & Soc. Phil. **11**, 1 (2016)

[27] Aloni, M.: Logic and conversation: the case of free choice. Semantics and Pragmatics **15**, 5–1 (2022)

[28] Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial intelligence **44**(1-2), 167–207 (1990)

[29] Blackburn, P., Rijke, M.d., Venema, Y.: Modal Logic. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, UK (2001)

[30] Belnap, N., Perloff, M., Xu, M.: Facing the Future: Agents and Choices in Our Indeterminist World. Oxford University Press, ??? (2001)

[31] Boudou, J., Lorini, E.: Concurrent game structures for temporal stit logic. In: 17th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), pp. 381–389 (2018). ACM; International Foundation for Autonomous Agents and MultiAgent Systems ...

[32] Velleman, J.D.: How to share an intention. Philosophy and Phenomenological Research: A Quarterly Journal, 29–50 (1997)

[33] Finnis, J.: Intention and Side-effects. Faculty of Law, University of Toronto, Toronto,Canada (1988)

[34] Knobe, J.: The concept of intentional action: A case study in the uses of folk psychology. Philosophical studies **130**, 203–231 (2006)

[35] Knobe, J.: Intentional action and side effects in ordinary language. Analysis **63**(3), 190–194 (2003)

[36] Khaitovich, D., Özgün, A.: Hyperintensional intention. In: Bjorndahl, A. (ed.) Proceedings of TARK 2025. EPTCS, pp. 44–60. Open Publishing Association, Waterloo, Australia (2025). To appear

[37] Forrester, J.W.: Gentle murder, or the adverbial samaritan. The Journal of Philosophy **81**(4), 193–197 (1984)

[38] Fagin, R., Halpern, J.Y.: Belief, awareness, and limited reasoning. Artificial intelligence **34**(1), 39–76 (1987)

[39] Linder, B., Hoek, W., Meyer, J.-J.C.: Formalising motivational attitudes of agents: On preferences, goals and commitments. In: Intelligent Agents II Agent Theories, Architectures, and Languages: IJCAI'95 Workshop (ATAL) Montréal, Canada, August 19–20, 1995 Proceedings 2, pp. 17–32 (1996). Springer

[40] Gillon, R.: Foreseeing is not necessarily the same as intending. BMJ: British Medical Journal **318**(7196), 1431 (1999)

[41] Bratman, M.E.: Shared Agency: A Planning Theory of Acting Together. Oxford University Press, Oxford, UK (2013)