University Utrecht

Faculty of Humanities

Department of Information and Computing Sciences

Artificial Intelligence

Deviant Causal Chains in XSTIT Logic

Thesis for the Degree of Bachelor of Science

Author:
A.N. Hermans

Supervisor:
Dr. Ir. J.M. Broersen

July 5, 2013

Contents

1 Introduction **2**

2 Deviant Causal Chains **4**
2.1 Causal theories of action 4
2.2 Definition of deviant causal chain 6
2.3 When and why deviant causal chains were introduced 7
2.4 A problem for the causal theory of action? 9
2.5 Summary 14

3 Modelling in XSTIT Logic **16**
3.1 Introduction to STIT logic 16
3.2 The successful sniper example 17

4 Deviant causal chains in XSTIT **20**
4.1 Two competing theories 20
4.2 The sniper example 21
4.3 Introduction to SPASS 22
4.4 Possible representations 23
4.5 Is it possible? 29

5 Conclusion **31**

Appendices **36**

A BI-frame **36**

B KI-frame **40**

1 Introduction

Artificial Intelligence is an interdisciplinary study of which philosophy and logic are two main areas. This thesis tries to combine both areas in a way which is not straightforward. In particular, it examines deviant causal chains in combination with the XSTIT logic of Broersen [11]. XSTIT logic is a STIT logic which can express properties of agency such as an agent making a choice or performing an action. A better understanding and ability to model agents and their choices can be useful for understanding human behavior and intelligence, for which Artificial Intelligence's ultimate goal is to design artificial intelligent computer systems and software. While logical models of agency are important, we should not underestimate the scientific role of philosophy within Artificial Intelligence. Debating about what an action precisely is, how actions are caused by intentions and how this causal chain might be deviant, are topics examined by the philosophy of action. This knowledge can provide a basic framework on which the logics of action can elaborate. In this thesis, both philosophy and logic are examined in a combination of literature research with practical research. The practical research consists of modeling XSTIT frames and working with the theorem prover SPASS [43].

The main question this research is trying to answer is the following: 'Can deviant causal chain examples be modeled in XSTIT logic?'. This question encompasses the broader question: 'Can causal relations be expressed in a form of STIT logic?'. If these questions will be positively answered, we have obtained a union between two different theories within Artificial Intelligence.

Combining deviant causal chains with STIT logic is a challenging task. The theory behind deviant causal chains is based on an ontological paradigm while the theory behind STIT logic is based on a modal paradigm. On the one hand, the ontological paradigm behind the causal theory of action, supports the view that actions are events, that there is a causal link between an action and an intention and that events are deterministically caused by other events. Davidson is one of the main supporters of this paradigm. One of the differences between this paradigm and the theory behind STIT logic is that STIT logic elaborates on the idea that events are not deterministically caused. Moreover, STIT logic focuses on agency, agents and the consequences of their choices, without taking into account *how* an agent is bringing about a specific action in terms of the causal relations involved. Thus, while deviant causal chains are all about causality, STIT logic does not explicitly express causality in its logical framework. That is why the attempt of modeling deviant causal chains in XSTIT logic is closely connected to the broader question if causal relationships can be expressed in STIT logic. And that is also exactly why this research is challenging and interesting.

In order to answer the main question, we have to be acquainted with deviant causal chains first. Section 2 will therefore examine related questions about deviant causal chains. Firstly, it will provide background information about the philosophy of action and intention and will identify key players in the debate. Secondly, this section will examine what a deviant causal chain exactly is and when and why this term was introduced. Then the question to what extent deviant causal chains are considered to be a problem for the causal theory of action will be discussed and various arguments will be examined. Finally, this section will conclude whether the solutions provided are adequate enough to safe the causal theory of action from the problem of causal deviancy.

When the philosophical background of deviant causal chains will be provided, section 3 will introduce us to STIT logic and the extended XSTIT logic of Broersen. To get acquainted with modeling in this form of logic, this section will model an example in a KI-extended XSTIT frame.

The last section will examine whether deviant causal chains can be modeled in XSTIT logic. Firstly, more elaboration on why exactly the main question of this research is challenging will be provided. The following subsection will show a XSTIT frame in which a deviant causal chain example might be modeled. This first attempt will give us a feeling about what kind of frame might be suitable and what other properties of both deviant causal chains and the logical framework we have to keep in mind. Then, with the help of theorem prover SPASS, logical properties of the logic will be investigated and related to the deviant causal chain problem. Finally, with an examination of possible representations and derivations, a conclusion will be drawn whether or not the attempt of modeling deviant causal chains in XSTIT logic was successful.

I want to thank Jan Broersen for his help and guiding role during this research. He was a great inspiration and offered me the freedom to explore the research question myself. Many suggestions and helpful hints as well as explanation and important knowledge have been provided by him for which I want to thank him too.

2 Deviant Causal Chains

Deviant or wayward causal chains examples were introduced as counterexamples to causal theories of intention; theories which advocate that, in order for an agent to act intentionally, it is both necessary and sufficient that an agent has a mental state of its intended action and that the agent being in that state causes him to achieve his target [36]. This condition can be stated as follows:

> "A subject intentionally ϕ-s if and only if their intending to ϕ causes their ϕ-ing." [36, p.1]

Cases of causal deviancy demonstrate that the causal condition (intending to do ϕ causes the action of ϕ-ing) is not sufficient for an action being intentional. They give rise to the problem how something unforeseen can or cannot be represented in the agent's initial intention or plan [22]. Some philosophers argue that deviant causal chain examples are no problem for causal theories of intention [28, 39]. On these arguments I will elaborate later in this section. The first subsection will introduce the concept intention, causal theories about intention and key figures in the debate about intention. Secondly, it will be described what a deviant causal chain exactly is, why and how deviant causal chains were introduced and why they might be a problem for causal theories of intention. The last section will also investigate some suggested solutions.

2.1 Causal theories of action

What is exactly an intention? If someone has the intention to raise his arm, but never raises his arm, how can we then describe this intention? What is the relation between a reason and an action? When can we say that someone acts intentionally? To what extent does an intention *cause* an action? If I have the intention to go to the doctor, but if I know that going to the doctor means having pain, do I then intentionally have pain?

These questions are just a few examples of questions concerning the philosophy of action, causality and intention. The debate about to what extent an intention *causes* an action plays a central role in this area. To understand this discussion we first have to understand what an action exactly is. With defining the concept of action, the problem arises where to draw the line when someone is performing an action. It seems that an action is directed to achieve a certain goal. We can for example passively cough, shiver or scratch ourselves, but although these events are all executed by us, we do not really consider these events as actions. Solving whether something is just a movement or really an action is a question Wittgenstein has proposed by asking: "What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?" [46, p.161]. Just an arm going up can be defined as an 'event', which is something that merely happens to an entity [23]. An important question is whether actions can be considered as kinds of events or not. The common view is that an action is an event, but a special sort of event; it is an event carried out by an agent.

Events are caused by other events. Actions however, are not caused by events but by the agent himself [15, 38]. It is essential for an action to have an 'owner'; that there is always an agent performing the action. For many years philosophers debated intensely about whether or not actions can be causally explained by the reasons of an agent. Davidson, who plays an influential role in this debate, states that the "reason rationalizes the action" [16, p.3];. According to his view, an agent's action can be causally explained by the reasons why he did so. More precisely,

Davidson emphasizes that an action can be explained by an agent's "primary reason" which he defines as the agent's "pro attitude" and its "related belief" [16, p.4]. Davidson elaborates on this view in several essays, on which I will not elaborate. However, consider the following example to understand Davidson's general idea:

> A primary reason of me taking my medicine might be that I want to get better (pro-attitude) and I belief that taking my medicine leads to making me better (belief).[1]

In this example me taking my medicine can be causally explained by my pro-attitude and my belief. Davidson notes that the explanation of a primary reason being the cause for an action does not have to be as strict and deterministic as the explanation in causal laws such as in natural sciences. However, it still holds that the primary reason can be causally responsible for the action [16, p.4-19]. Davidson argues that action is "intentional under some description", so that an action requires a formulated intention [16, p.50]. This also requires that the agent himself knows what he is doing under some description. To use the example above, suppose I take my medicine, I intend to do whatever is needed to take it and I know that I am doing so. The pro-attitude together with the belief rationalize an action if they cause it "in the right way" [16, p.79]. Other philosophers such as Goldman [21] support that a reason or intention can be causally linked to an action.

Not everyone agrees with Davidson. It has also been argued that there is not a causal relation between a reason and an action. The neo-Wittgenstein movement argues that reasons cannot causally explain actions because of conceptual reasons (one of these philosophers is Taylor in [38]). Other philosophers, such as Anscombe, partly agree. Anscombe agrees with Davidson that an action can be intentional under some description but not under *every* description [1]. For example, I might be observing someone who is moving his arm up and down while holding a handle. Suppose this man is moving his arm up and down exactly on the rhythm of the music which is playing. A good way to determine this man's intention is to ask why he is doing that. If we ask 'Why are you moving your arms on the rhythm of the music?' and the man says 'I was not aware of that!' then clearly moving his arms on the rhythm of the music was not his intention. The action of moving his arm up and down is not intentional under this description or other descriptions such as "contracting muscles", because that's not why this man is moving his arm up and down. In this case the action is intentional under the description of "pumping water" (because the man is actually moving his arm up and down because he wants to pump water) [1, p.40]. Anscombe's view influenced Davidson's theories about intention and he expended his theories on her claims. However, Anscombe never accepted Davidson's causal theory of action. She argues that there is a distinction between a cause and a reason [1]. Other philosophers such as Wilson [45] and Ginet [20] agree with Anscombe. Another important argument of Anscombe is that having an intention is a necessary but not a sufficient condition for an intentional action [1].

Also Mele [26], Bratman [9] and Searle [34] argue that merely having an intention is not enough for an act to be intentional. They argue against what has been called 'the Simple View' [9]. Searle introduced the distinction between prior intention and intention in action [34]. With a prior intention "the agent acts on his intention", while the intention in action is "just the Intentional content of the action" [34, p.84]. Someone can do something intentionally without having a prior intention, thus a prior intention is not a necessary condition for an intentional action [34].

We have examined a few important theories related to intention. Much more has been argued about action, causality and intention, on which will not be further elaborated. Important to

[1] This is my own example based on examples from Davidson in [16].

note is that the question if intention is causally related to action and whether or not intention is a sufficient and/or a necessary condition for an intentional action is where the deviant causal chain examples play a role. As mentioned earlier, Davidson accentuates that a pro-attitude and a belief must cause the action in the 'right way' in order for the action being intentional. If they cause the action in the 'wrong way', it might be that the action is not intentional any more. But what can be considered as the 'right way'? And to what extent can these examples serve as counterexamples against the causal theory of action? Before we answer this question, the next section elaborates on the definition of a deviant causal chain.

2.2 Definition of deviant causal chain

Let us try to understand what a deviant causal chain exactly is by discussing some examples:

> "A certain main desires to inherit a fortune; he believes that, if he kills his uncle, then he will inherit a fortune; and this belief and this desire agitate him so severely that he drives excessively fast, with the result that he accidentally runs over and kills a pedestrian who, unknown to the nephew, was none other than the uncle." [15, p.30]

> "A man may try to kill someone by shooting at him. Suppose the killer misses his victim by a mile, but the shot stampedes a herd of wild pigs that trample the intended victim to death." [16, p.78][2]

In the last example the man's intention to kill his victim rationalizes and causes him to shoot his victim in order to kill him. This shot causes the stampeding of the herd of wild pigs, which leads to the death of the man's victim. Thus the man's intention *causes* his victim's death, yet we would not say that the man killed his victim intentionally. Davidson, who is referring to these kind of examples as "external causal chains", puts forward another example where an "internal causal chain" is involved:

> "A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally." [16, p.79][3]

In all these examples, some "control-undermining" state or event takes place between the agent's "reason states" and the event produced by that agent [33, p.188]. However, there is a difference between the last example and the other two examples. The last example is a case of "basic deviance", while the first two examples are cases of "non-basic deviance" [22, 7, 33].[4] Basic deviance effects the causality between the mental state and the basic action, such as in the climber example. Here the control-undermining state takes place between the reason state and the basic action (the action for which the agent does not need to undertake any further steps) [33]. It is discussable whether or not the climber really performs an *action*, but if we assume he does, the climber's desire to loosen his hold on the rope is here the 'reason state', the action of letting go of the rope is the 'basic action' and the state of nervousness is the 'control-undermining state' which

[2] I will refer to this example as 'the sniper example'.

[3] I will refer to this example as 'the climber example'

[4] Brand makes the distinction between "antecedential" and "consequential waywardness" [8], Davidson between "internal" and "external" deviant causal chains [16] and Mele between "primary"and "secondary" deviancy [25]. The third category which Bishop and Enç distinguish, "second-agent" deviancy, I will not consider because it can be argued that this deviancy is a special case of either basic or non-basic deviance [18].

occurs in between. Non-basic deviance effects the causality between the basic action and the non-basic action [33]. A non-basic action is an action for which more steps have to be undertaken in order to perform the action. In this case the action of killing is preceded by more actions (for example the action of shooting). The control-undermining state occurs between the basic action and the eventual outcome which the agent wants to achieve by performing this action, in this case between the agent firing the gun and the death of his victim [33].

A deviant causal chain is a causal chain which deviates from the 'normal' or 'right' causal chain. But what is this 'normal' or 'right' causal chain? In all deviant causal chain examples an agent has the intention to do something and by having this intention the agent causes a deviant causal chain to take place which causes the intended action to happen, but yet we would not classify the action as intentional. We can consider the agent's plan as the 'right sort of way' and the deviant causal chain as a chain which was not embedded in his plan.

Another definition of causal deviancy was introduced by Peacocke [31]. Peacocke defines the difference between a normal and a deviant causal chain in the way in which they can explain an action. In the 'normal' case a causal explanation will succeed as what Peacocke calls a "differential explanation", which means that we can describe the intentions of the agent and the resulted action in a way that they are functional dependent on each other [29, p.106]. Note that Peacocke accentuates that this functional relationship applies only under their physical descriptions [6]. In the deviant case a causal explanation will fail as a differential explanation because the cause and the resulted action cannot be described as the former a function of the latter [29].

Although we are temped to conclude that in these examples the agent in question did not perform the act intentionally, it actually depends on how the action is defined. If in the sniper example, the agent's intention was to kill his victim no matter what, he succeeded. However, if he intended to kill the other agent by shooting him, he did not succeed [22]. That the sniper did not intentionally kill will be the most likely given the fact that probably the sniper did not intend to startle the wild pigs, maybe he did not even notice them. And even if he did notice the wild pigs and might intentionally have startled them, how could he have known that shooting would lead to the pigs trampling his victim to death?

Now that we have seen how deviant causal chains already give rise to some interesting questions, it is time to find out when and why they exactly were introduced.

2.3 When and why deviant causal chains were introduced

The first example of causal deviancy was introduced by Chisholm [15] when he examines in detail the notion of an 'act'. He stresses that the relation between what we want or desire and what we do is not as simple as some philosophers assume when they attempt to define purpose "in terms of belief, causation, and desire" [15, p.29]. Although a purpose is comparable to a desire because it is involved in an act, it is not the same as a desire. Chisholm particularly attacks Ducasse's definition of a purposive act in terms of beliefs and desires. This definition does Ducasse introduce as follows:

> "To be able properly to speak of an act (or event) as purposive ..., what is essential ... is that the following elements be present, or be supposed, by the speaker, to be present:
> 1. *Belief* by the performer of the act in a law (of either type), e.g., that If X occurs, Y occurs.
> 2. *Desire* by the performer that Y shall occur.
> 3. *Causation by that desire and that belief jointly*, of the performance of X."
> [17, p.543]

Chisholm made the following objection to this definition:

> "Suppose, for example,
> (i) a certain main desires to inherit a fortune;
> (ii) he believes that, if he kills his uncle, then he will inherit a fortune; and
> (iii) this belief and desire agitate him so severely that he drives excessively fast, with the result that he accidentally runs over and kills a pedestrian who, unknown to the nephew, was none other than the uncle. The proposed definition of purpose would require us to say, incorrectly, that the nephew killed the uncle in order to inherit the fortune." [15, p.30]

This objection reveals that Ducasse's definition of a purposeful act allows unintentional actions to be qualified as intentional and that therefore the definition is not accurate. Chisholm introduces another definition of an action which emphasizes that an agent should be a causal factor of an act.

Philosophical literature often remarks that Davidson is the one who brought deviant causal chains to our attention, especially as a problem for causal theories of action (for example Tännsjö in [37]). Also it is noted that Davidson's approach, the view that we are not able to describe *exactly* how beliefs and desires 'normally' cause intentional actions, dominated the debate for a long time [37]. In his essay 'Freedom to Act' Davidson argues that the causal theory of action should not be abandoned despite of its difficulties [16, p.64]. The causal theory of action, as mentioned, encompasses the view that actions are caused by states such as beliefs and desires.

One well-known attack for the causal theory of action is that if actions are caused by certain states such as beliefs and desires, then the freedom to act must be a "causal power" [16, p.63]. Davidson remarks that so far "no proposed account meets all objections" [16, p.63]. With his analysis of the notion of an action, Davidson examines attempts made of formulating the following laws all needed for a proper causal theory of action: "a law stating conditions under which agents perform intentional actions; an analysis of freedom to act that makes it a causal power; and a causal analysis of intentional action " [16, p.76]. At this point Davidson refers to Armstrong who describes the difficulty in specifying the conditions for an action to be intentional. That difficulty is the deviant causal chain, the difficulty that "the attempt may bring about the desired effect in unexpected or undesired ways" [16, p.78]. This is where Davidson mentions the examples of non-basic deviancy and introduces the climber example as a form of basic deviancy.

The causal theory of action is threatened by the existence of deviant causal chains. But to what extent is causal deviancy a real problem for the causal theory of action? The next section tries to answer this question.

2.4 A problem for the causal theory of action?

The following quote clearly identifies for what kind of action theories the deviant causal chain examples are considered to be a problem:

> "The problem of deviant causal chains is endemic to any theory of action that makes definitional or explanatory use of a causal connection between an agent's beliefs and pro-attitudes and his bodily movements. Other causal theories of intentional phenomena are similarly plagued." [22, p.69]

As soon as we speak about a causal connection between our beliefs and desires and our actions, deviant causal chains are involved because they can be part of this causal connection. Any theory making definitional or explanatory use of this causal connection, should account for the fact that these deviant causal chains can cause the same event, but in a different way. This is especially important for the causal theory of action. The main view within the causal theory of action is that actions are events. An event is an action if and only if it is caused and rationalized by the agent's mental states [33], mental states such as beliefs and desires. The problem with basic deviance is that the event in question is actually caused and rationalized by the agent's beliefs and desires. Yet we can argue that the agent did not perform an action; that the climber in the climber example did not perform an action. The loosening of his hold on the rope was an event which merely happened to him due to his nervousness [25]. In the same way we might consider the killing of the victim in the sniper example as not the sniper's action. Also Brand [8], Searle [34] and Thalberg [39] observe that in many cases of causal deviancy the outcome is not an action. Apart from the fact that the conditions set by the causal theory of action are not right, if the sniper would say 'I did not mean to do it like that', a causal theory of action should provide another explanation for the death of the uncle in the climber example [22]. A causal theory of action should be able to distinguish between right and deviant causal chains.

Examples of causal deviancy are also counterexamples to the definition of intentional action where beliefs and desires' rationalization and causation are both necessary and sufficient conditions for an act being intentional. Suppose we argue that the climber in the climber example did perform an action. Still the conditions for an act being intentional are met and yet we would not classify this action as intentional, or that the produced effect is intentional [22]. The same is true for cases of non-basic deviancy. It seems that beliefs and desires' rationalization and causation are a necessary, but not a sufficient condition for an act to be intentional.

At first hand, we may think that we only need to add the condition that the agent's intention should cause his action in a non deviant way [7]. However, a further specification of this casually non deviant way is needed without referring to "actions, agent-causation, exercises of control and the like" because otherwise the theory would be self-referential [7, p.98]. For example if we would state that in a proposition such as "intending to do ϕ causes ϕ-ing" ϕ should describe an intentional action, so in the case of the climber example ϕ stands for 'letting go of the rope intentionally', we obtain a "circular" and "self-referential" condition [36, p.2]. Although this might seem as a solution because the climber did let go of the rope accidentally, it does not bring us any further.

What it all comes down to is, as clearly described by Davidson, that if we somehow describe an intentional effect as caused by desires, beliefs or intentions then the problem is that not every causal connection between these attitudes and the effect is sufficient to consider the produced effect as intentional [16]. The causal chain here involved "must follow the right sort of route"

[16, p.78]. But what is this 'right sort of route'? And what is this 'right sort of causal connection'? The problem of causal deviancy remains specifying the 'right way'. We would have to formulate the exact conditions, which would have to be stricter then they are now, of when an act is intentional so that deviant causal chains are excluded. It has been intensively debated if this is possible. Let us examine some suggested solutions.

The common approach to deviant causal chains examples is what goes wrong is that the intended action did not happen according to plan [21, 8, 7, 27]. So the climber did not intentionally let go of the rope because the causal chain differs with his original plan. The time of letting go is different and/or the events involved in his plan [36]. However, regarding the time, an objection might be that one does not usually include an exact time within his or her plan [36]. And even if this would be the case, the deviant causal chain can also occur within the time we form our intention and our planned time of action while still bringing about the same action at that same time.

Maybe the specification that the events involved in the agent's plan were not exactly the same as the events happened, can be a solution. However, if we examine our plans closely, it is actually the case that every course of events is slightly different from the way in which we planned them to happen, but not every deviancy is a deviant causal chain. It is very rare that the things we plan happen exactly according to our plan because "the world is *thicker*, or denser, than our mental representations of it" [22, p.81]. Davidson reinforces this argument by stressing that there are endless ways in which an intention can result in an unintended effect. It seems that an intention just cannot specify all the characteristics which are needed for the act to be qualified as intentional [16]. It is impossible for us to include or exclude every possible causal chain in our intention. Other philosophers who agree with this argument are Morton [30] and Beck [2]. As Morton underpins, there is almost always another action which fits, besides the particular intended action, exactly within the specified intention [30]. So the question remains if we can make the causal theory strict enough without excluding too many cases [22].

In cases of non-basic causal deviancy different comparable solutions have been proposed that the intention must persist into the action in order for an act to be intentional. Where it is necessary to take further steps in reaching the intended action (as to shoot in order to kill) we must also take the further steps intentionally (and stampeding the wild pigs was not intentionally done). The agent must take the appropriate steps but must also trust the causal chains running through them. Schlosser formulates this as the importance of "the guiding role of the contents of mental attitudes" [33, p.190]. Searle states it slightly different that there must be a "continuous efficacy of Intentional content under its Intentional aspects" in order for an act to be intentional [34, p.138]. The intention must not only exists prior to the action, but must continue into the action. In addition, Mitchell agrees that the intention must persist into the time of acting. Keil puts forward the question how an agent can insure this efficacy. He objects that the agent cannot change anything once the causal chain has left the agent's body [22].

In cases of basic deviancy we cannot refer to an agent's action plan [33]. Consider the case of the climber example. The climber does not have a certain plan on how to loosen his hold on the rope, because loosening his hold on the rope is a basic act, there are no further steps required and the climber does not have to know how to bring about this act, he just does. Apparently cases of basic deviancy require another solution. Brand, Thalberg and Mele have proposed that an action should be *proximately* or *directly* caused by an intention [8, 27, 39]. This strategy has been called the "causal immediacy strategy" [7, p.138].

Mele advocates that an action is intentional if it is caused by "proximal intentions" directly [27, p.54]. Similarly, Brand allows no causal gap between the reason state and the bodily movement of an agent. The reason state must be "the beginning of the physiological chain" [8, p.20] without room for intervention. Brand uses the same term of proximate causation and supports that a mental state should proximately cause the agent's action in order for an act to be intentional. This means that there should not be a mediating link between the mental state and its effect. It should be the case that, for example in the climber case, the proximal intention initiates the sending of a signal to the motor cortex, which will lead eventually to sending the signals to the right muscles which will make the muscles move [8]. If this would happen, then the act would be intentional. However, what the mental state actually causes in the climber example is not an action, so this example is, according to this strategy, not a counterexample for the causal theory of action.

Also according to Thalberg, the examples of causal deviancy are ineffective as counterexamples [39]. He likewise supports the causal immediacy strategy. Thalberg elaborates on Frankfurt's notion that a movement has to be guided by a person to be an intentional action [19]. A movement has to be continuously caused by an intention [39]. Thalberg's view is comparable to Shaffer's argument that an agent must bring about the steps intentionally in order to bring about the end intentionally [35] and also to Mitchell's condition that the intention "must have persisted into the time when one acted, for the act to be intentional" [28, p.353]. In this way it can be argued that in the climber case, the climber did not act intentionally because the intention of letting go did not persist into the moment when he let go.

Thus, according to some philosophers, causal deviancy is no problem for the causal theory of action. Thalberg examines in his essay several deviant causal chain examples [39]. He remarks that they have forced further specification of the causal theory which has resulted in this theory being less "vague" and "vulnerable" [39, p.259]. So he actually observes a positive effect of the deviant causal chains for the causal theory of action. He describes the value of them as follows:

> "They force a causal analyst to be explicit: to specify that intentional Xing is produced by an intention, not just a desire and belief; that there should be minimal disparity between what you intend to do and what you end up doing; that your intention itself - not only some mediating occurrence - must bring about your Xing; and that your Xing should be an unproblematical instance of action." [39, p.259]

According to Thalberg there is no action which meets these specified conditions, which can be qualified as not intentional.

The causal immediacy strategy has not been accepted by everyone. It has been argued that the proximity solution is not adequate enough, that it does not advocate that the action is a *response* to a reason state. Bishop and Peacocke have pointed out that "an action that is done *for* a reason is a *response* to that reason" [33, p.191]. In the climber case for example, the causal chain seems to be deviant because it is *only* caused by the agent's desire to loosen his hold on the rope and that the action is not a response to this desire. The action is a response to the nervousness caused by the desire, but not to the desire itself. Philosophers who accept the proximity solution as a valid solution do not ought the way of rationalization relevant enough [33]. Another objection to the proximity solution is that reason states never can be proximate causes of action [7]. Moreover, even if the mental state is causally proximate to the beginning of the physiological chain, the physiological chain can still proceed via the deviant causal chain

(the nervousness state in the climber example) [7]. Bishop accentuates furthermore that the proximity solution would only be a valid solution if the types of physiological chains which are considered as causing an action, would be specified [7].

Schlosser's argument is comparable with Bishop and Peacocke's argument that an action should be a *response* to the agent's reason. He accentuates that it should be required that "reason states cause and casually explain actions in virtue of their contents" [33, p.192]. In other words, an action is not just a response to a reason, but it is a response to the content of the mental state. This condition is already proposed as a solution for non-basic deviance where the action should be guided by the reason state. Schlosser concludes that this solution can also be applied to basic deviance. Although in cases of basic deviance there is no plan involved, still the basic act is guided by the reason state because the reason state causes and causally explains the basic act in virtue of its intentional content [33]. In the climber example the movement is not a response to the intentional content of the nervousness state. We can say that the state of nervousness the climber is in, causes the movement of the climber in some way, but it does not cause so in virtue of its intentional content.

Tännsjö similarly holds that an action should be *responsive* to the content of the mental state [37], but he takes a step further than that. He provides a solution which is, according to him, enough to defend the causal theory of action. The solution he proposes is that each type of action possesses its own properties of when it is deviant or not [37]. In the climber case for example, the loosening of the agent's hands should be under the agent's control. Given each token of a specific action type, we are able to decide if it is really intentional or not [37]. Causal deviancy is no problem for the causal theory of action in this way.

Another solution has been proposed to leave the specification of the causal chain to the relevant special sciences such as neuropsychology. This solution has been called "Gricean Deference" with an analogy to Grice who suggested to leave the causal theory of perception to the appropriate special sciences [22, p.74-75]. Goldman is comparing intentional action with perception [21]. As not any causal connection between intention and action will be sufficient for an act to be intentional, also not every causal connection between the object and the "sensory content of the percipient" will be sufficient for the percipient to actually perceive the object [21, p.63]. However, Goldman accentuates that it is not fair to ask from a philosopher to specify this causal process, because there is specialized research necessary for obtaining this knowledge:

> "A complete explanation of how wants and beliefs lead to intentional action would require extensive neuropsychological information, and I do not think it is fair to demand of philosophical analysis that it provides this information." [21, p.62]

Although Goldman assures that we have a certain 'feeling' to differentiate between deviant and non-deviant causal chains, he leaves the specification of these processes to the special sciences. Armstrong and Mele are other philosophers who defend a form of Gricean Deference. Armstrong reinforces Goldman's argument by comparing our "mechanisms" with mechanisms in a computer which may, such as in a computer, "malfunction" [22, p.21]. Mele underpins that in the climber example we can psychologically identify an anomaly in the motor control system of the climber [25]. He suggests that we might be able to design a human in such a way that no such malfunction can occur. Although he leaves it at question if this would be possible, if we would be able to design such a human, the climber example would not be a problem any more.

Critique on the Gricean Deference strategy is that this strategy underestimates the problem of actually identifying a deviant causal chain [22]. The 'right' sort of causal chain does not possess any distinct physical properties to differentiate it from the 'wrong' one. Keil is convinced that the distinction between right or wrong is not one which neuro-sciences or other sciences can make [22]. It can be considered as a philosophical task to classify the reason states, the event produced by that agent and the control-undermining state which produced the effect together. Moreover, if Gricean Deference would be a solution, it would only work for the cases of basic deviance.

Bishop believes that, after examining several solutions for both basic and non-basic deviance, he has a solution compatible with the causal theory of action. He makes the following statement:

> "I believe that I am now in a position to make a well-founded claim that basic intentional action may be analyzed as matching behavior that is sensitively caused by the agent's basic intention, and in a context where any feedback to central mental processes returns to the agent's, rather than to anyone else's brain. This analysis offers what is needed for a defense of CTA (Causal Theory of Action) because it provides necessary and sufficient conditions for basic intentional action that make no essential reference to agent-causation." [7, p.171]

Thus, the sensitivity strategy mentioned earlier seems a promising solution. With the last condition mentioned by Bishop, cases of "preemptive heteromesy" are excluded where the deviant causal chain goes via the intention of another agent [7, p.157]. At the same time the conditions Bishop has formulated allow causal chains to take place where the help of another agent is consciously used. Finally, this definition is not too strong in excluding cases where intentional basic actions take place with feedback to the "brain's central processes" [7, p.171].

Can we conclude that the sensitivity solution is the best solution proposed? Unfortunately, it is too early to conclude that, since several solutions had to be left unexplained and not all solutions we have examined, have been evaluated in full detail. The debate about causal deviancy is too complex to fully explore it in the space we have here. For every solution we can find arguments for and against it. It seems like the sensitivity strategy is a really adequate solution. However, against the sensitivity strategy, it has been argued that it only focuses on the *initiation* of the agent's bodily movements by its mental state, while much more is involved in the realization of the agent's control over his actions [24]. The sensitivity strategy does not take into account possible corrections and adaptations of the agent's intention to the environment, while this is also an essential element of agency [24]. The "sustaining causation strategy" which does deal with this element of agent control, has been most accepted among philosophers according to Mayr [24, p.121].

Deviant causal chains are a problem for the causal theory of action. However, with all the different solutions suggested and examined, I would conclude not to abandon the causal theory of action despite of its difficulties. For every solution there are pro and cons to debate about, but in the end there are reasonable enough philosophers who have argued that their solution safes the causal theory of action, or that causal deviancy is no problem at all for the causal theory of action. Now it is the matter of deciding which solution is really the best, or which solutions have to be combined to obtain the perfect solution. At least it seems that this can be done in order to obtain a causal theory of action which can handle causal deviancy.

2.5 Summary

The causal theory of action encompasses the view that actions are events and that an action is intentional if and only if it is rationalized and caused by the agent's beliefs and desires. To what extent an intention really *causes* an action is one of the main discussions within the philosophy of action and intention. The first case of causal deviancy was introduced by Chisholm and brought to more attention by Davidson who also introduced different kinds of deviancy: basis and non-basic deviancy. In all the cases of causal deviancy some control-undermining state or event takes place between the agent's reason states and the event produced by that agent. Basic deviance effects the causality between a mental state and a basic action, while non-basic deviance effects the causality between a basic action and a non-basic action.

Deviant causal chain examples were introduced as counterexamples to the causal theory of action. They demonstrate that the rationalization and causation of an agent's intention is not a sufficient condition for an act to be intentional. New conditions should be introduced without referring to notions such as action and agent-causation to avoid a self-referential theory. Different solutions have been proposed in order to safe the causal theory of action. By no means did this section examine every proposed solution, or did it examine the proposed solutions in full detail, because that would require too much space. However, this section did summarize the main solutions and did examine some important arguments for and against them. In this way we are now acquainted with the debate and its key arguments so that we can think about this subject ourselves.

There are solutions which at first hand do not seem suitable. That an intentional action should happen exactly according to plan is a problem because it is impossible to include or exclude every causal chain in an intention or plan. If an agent should take the steps needed to achieve the non-basic action intentionally, the question is how an agent can ensure this. This solution will anyhow not work for cases of basic deviancy. Here the causal immediacy strategy seems a solution where there is no causal gap allowed between the intention and the bodily movement of the agent. But an objection to this solution is that even in cases of basic deviancy, still the physiological chain can proceed deviantly after the intention was formed. Another objection that has been put forward is that the physiological chains should be further specified.

This further specification of physiological chains might be left to the neurosciences, a strategy called Gricean Deference which is defended by Goldman, Armstrong and Mele, although in different variations. A problem with this strategy remains that it is probably not a neuroscientific task to differentiate between 'right' and 'deviant' causal chains because they do not possess distinct physical properties. Thus it still remains a philosophical task to distinguish between intentional and non-intentional actions.

To sum up, every solution has been in some way attacked and undermined. However, until now, I find the sensitivity solution of Bishop and Peacocke and also defended in slightly different versions by Schlosser and Tännsjö the most appealing. This solution requires that the action should be a response to the reason state or in other words, that the action is sensitively caused by the agent's intention. In his essay, Bishop [7] clearly examines many arguments and defends that his solution excludes the cases of causal deviancy, without excluding the normal cases. The sensitivity strategy seems to provide the conditions for an act to be intentional in such a way, that there is no non-intentional action which satisfies the conditions or an intentional action which does not satisfy the conditions.

That there has also been critique offered against the sensitivity strategy, shows us that we cannot take any solution for granted. However I conclude, although deviant causal chains are a real problem for the causal theory of action, that the causal theory of action does not have to be abandoned despite of the difficulties the deviant causal chains have brought to our attention. I agree with Thalberg that causal deviancy has forced us to reconsider and improve the causal theory of action. Deviant causal chains are not only threatening to the causal theory of action, but are also beneficial for it. It might be the case that the 'perfect' solution is still not there, but it seems like we are heading to one.

3 Modelling in XSTIT Logic

3.1 Introduction to STIT logic

This thesis examines if it is possible to model deviant causal chain examples in the XSTIT logic of Broersen as described in [11]. This logic is a variation on STIT logic, which finds its origin in philosophy and was first proposed by Belnap [4]. STIT logic is a philosophical logic of agency and an application of temporal logic, a logic based on the branching time theory originally developed by Prior [32]. Here time is represented as a branching tree which has multiple branches in the future, but a single route to the past with non-determinism implicit in the branching tree structure [5]. As Belnap and Perloff describe, with this branching time tree structure, the choices agents can make can be represented as sets of possible futures passing through the particular choice point the agent is in [5]. This makes the logic suitable to describe properties of agency.

STIT is an acronym for 'Seeing To It That'. The expression in STIT logic $[\alpha \text{ stit} : Q]$, meaning agent α 'sees to it that' some proposition Q, is interpreted within the structure of branching time. To be more precise, the expression $[\alpha \text{ stit} : Q]$ is evaluated in a temporal branching structure where $[\alpha \text{ stit} : Q]$ means that Q is guaranteed by a prior choice of agent α, so it describes agent α carrying out an action [5]. Thus the sentence $[\alpha \text{ stit} : Q]$ is agentive for agent α. In most variations of STIT logic, $[\alpha \text{ stit} : Q]$ holds in a particular state if Q holds at that state in all the histories which are selected by the agent's choice function. An agent choosing between actions is defined as an agent determining which history will be among the actual histories.

The uniqueness of STIT logic is that it can express properties and aspects of agency (such that a choice is made or an action is executed by an agent), which cannot be expressed in dynamic logics [5, 11, 10]. STIT logic even has been called "the most suitable logical systems dealing with agency, both in terms of expressiveness and formal properties" [40, p.1]. Broersen's XSTIT logic distinguishes itself from other STIT logics because actions take effect in next states, which are defined as the immediate successors of the actual state [11, p.4]. Here $[A \text{ xstit}]\phi$ stands for "agents A jointly see to it that ϕ in the next state" [11, p.4]. Acting by a group of agents A is identified with "ensuring that a condition holds on all (dynamic) states that may result from exercising a choice" [11, p.6].

Broersen extends his XSTIT logic by adding a knowledge operator K_a and an intention operator I_a to the standard XSTIT frames in order to model the concept of "knowingly doing" and "intentionally doing" [11, p.11,14]. To model the notion of an unsuccessful choice, Broersen proposes the notion of "believingly doing" (with the belief operator B_a) as a variation on knowingly doing [11, p.22]. These operators seem very useful for modeling deviant causal chains because they make it possible to model the intention and knowledge of an agent. The following sections use the definitions and truth conditions of the KI- and BI-extended XSTIT frames as Broersen defines these. In this thesis, the frames are also visualized as Broersen visualizes them. For a full and detailed description of the logic, I refer to [11].

Note that the following sections require basic knowledge and understanding of modal logic. This thesis will not elaborate any further on the explanation of the different accessibility relations (reflexive, symmetric, transitive, serial and euclidean), nor the logical systems that emerge from these (K, D, T, S4, S5, etc.), nor explanation about frames, models, satisfiability and validity. For explanation about the different accessibility relations and logic systems I refer to [41].

3.2 The successful sniper example

Before the next section will investigate how to model one of the deviant causal chain examples, the sniper example, in a XSTIT frame, this section will first model an example without a deviant causal chain involved to get familiar with modeling in a XSTIT frame. The successful sniper example will be a modified version of the sniper example:

> A man kills somebody by shooting at him and succeeds in this way.

Which can be rewritten as:

> Agent a kills agent b by shooting agent b and succeeds in this way.

For this example the KI-extended XSTIT frame will be used, because the agent's choice is successful and it will be interesting to see the difference between a KI- and BI-extended frame (next section). The example will be further rewritten as:

> Agent a intentionally kills agent b by knowingly shooting agent b and succeeds in this way.

By modeling this example in a KI-extended XSTIT model M1, I want to preserve the information that agent A kills agent B *by shooting him*. The part 'by shooting agent b' is rewritten as 'by knowingly shooting agent b' and not as 'by intentionally shooting agent b' because what agent a knows to be doing can be possibly more than what it intends to be doing. In other words: agent a can knowingly perform different actions with the same intention. There are more ways in which agent a can intentionally kill agent b. To show this, I assume that agent a, being in static state s_1 has two possible static states to go to where it holds that he kills agent b: s_2 and s_3. The agent can be knowingly and causally doing something different, but intentionally the agent is doing the same. In this example agent a can kill agent b by choosing to fight with him or by choosing to shoot him. Because for an act to be intentional, the agent must have a choice of also not doing that particular act, it is also possible for agent a to choose to not kill agent b (static state s_4).

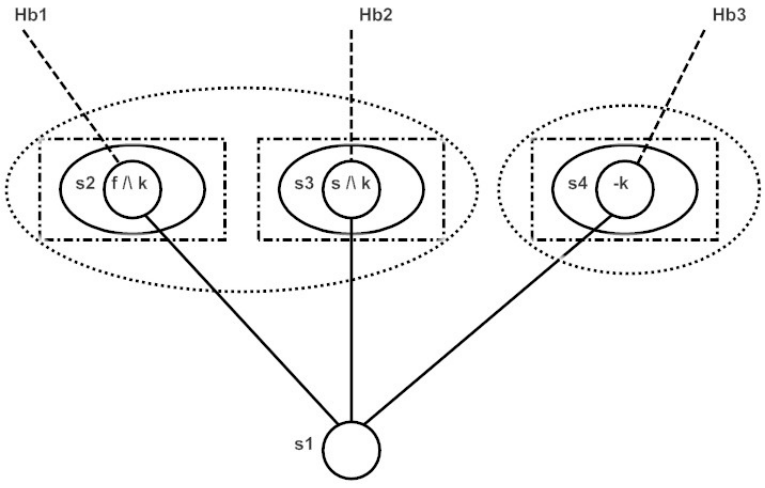

Figure 1: The successful sniper example in KI-extended XSTIT model M1.

With $h_1 \subseteq \{s_1, s_2\}, h_1 \in Hb_1$, $h_2 \subseteq \{s_1, s_3\}, h_2 \in Hb_2$, $h_3 \subseteq \{s_1, s_4\}, h_3 \in Hb_3$, f = 'fighting agent b', s = 'shooting agent b' and k = 'killing agent b'.

The atomic propositions which hold in a dynamic state are written in the static state on which the dynamic state is based on. Thus $f \wedge k$ written in static state s_2 means that it holds that $M1, \langle s_2, h_1 \rangle \models f \wedge k$.

Frame F1 visualizes the choices for agent a (relation $R_{\{a\}}$) as ellipses grouping the different possible sets of *next* dynamic states it can reach from the static state before. So in figure 1, the ellipse around static state s_2 groups the possible set of next dynamic states it can reach from static state s_1, in this case only $\langle s_2, h_1 \rangle$. The ellipse around static state s_3 groups the dynamic state $\langle s_3, h_2 \rangle$ and the ellipse around static state s_4 groups the dynamic state $\langle s_4, h_3 \rangle$. Note that a XSTIT-frame is defined as having an infinite set of static states (definition 2.2 in [11, p.5]) and the figure only shows a finite set of states. That is also why the histories written down (h_1, h_2, h_3) are actually not complete. However, since the other states and histories are not relevant now, I do not denote them.

So, in this case, we can only think of three histories, one running through history bundle Hb_1, one through Hb_2 and one through Hb_3. That is why in this case the ellipses all group only one dynamic state, while usually in more complex frames the ellipses are grouping more dynamic states together. The epistemic equivalent dynamic states are grouped together by the dotted rectangles and the intentional equivalent dynamic states are grouped together by the dotted ellipses. Static state s_1 is not surrounded by any ellipses or rectangles because the assumption is that agent a is currently in static state s_1.

The following holds in model M1 corresponding to frame F1:

$$\begin{array}{ll} M1, \langle s_2, h_1 \rangle \models & f \wedge k \\ M1, \langle s_3, h_2 \rangle \models & s \wedge k \\ M1, \langle s_4, h_3 \rangle \models & \neg k \\ M1, \langle s_1, h_1 \rangle \models & K_a[a \text{ xstit}](f \wedge k) \\ & I_a[a \text{ xstit}]k \\ M1, \langle s_1, h_2 \rangle \models & K_a[a \text{ xstit}](s \wedge k) \\ & I_a[a \text{ xstit}]k \\ M1, \langle s_1, h_3 \rangle \models & \neg I_a[a \text{ xstit}]k \end{array}$$

To get an idea of what the truth conditions are and how they work, I show how some of the propositions hold in this model. The truth conditions are the ones from [11, p.8,11]:

$$\begin{array}{rl} M, \langle s, h \rangle \models K_a\phi & \Leftrightarrow \langle s, h \rangle \sim_a \langle s', h' \rangle \text{ implies that } M, \langle s', h' \rangle \models \phi \\ M, \langle s, h \rangle \models [A \text{ xstit}]\phi & \Leftrightarrow \langle s, h \rangle R_A \langle s', h' \rangle \text{ implies that } M, \langle s', h' \rangle \models \phi \\ M, \langle s, h \rangle \models I_a\phi & \Leftrightarrow \langle s, h \rangle i_a \langle s', h' \rangle \text{ implies that } M, \langle s', h' \rangle \models \phi \end{array}$$

This is how $M1, \langle s_1, h_1 \rangle \models K_a[a \text{ xstit}](f \wedge k)$ and $M1, \langle s_1, h_2 \rangle \models K_a[a \text{ xstit}](s \wedge k)$ hold:

- $\langle s_1, h_1 \rangle \sim_a \langle s_1, h_1 \rangle$ because the dotted rectangle around s_2 in the visualized frame only groups the dynamic state based on h_1.
- $\langle s_1, h_1 \rangle \sim_a \langle s_1, h_1 \rangle$ and $M1, \langle s_1, h_1 \rangle \models [a \text{ xstit}]f \wedge k$, thus $M1, \langle s_1, h_1 \rangle \models K_a[a \text{ xstit}](f \wedge k)$.

- $M1, \langle s_1, h_1 \rangle \models [a \text{ xstit}](f \wedge k)$ because $\langle s_1, h_1 \rangle R_a \langle s_2, h_1 \rangle$ and $M1, \langle s_2, h_1 \rangle \models f \wedge k$.
- $\langle s_1, h_2 \rangle \sim_a \langle s_1, h_2 \rangle$ because the dotted rectangle in the visualized frame only groups the dynamic state based on h_2.
- $\langle s_1, h_2 \rangle \sim_a \langle s_1, h_2 \rangle$ and $M1, \langle s_1, h_2 \rangle \models [a \text{ xstit}](s \wedge k)$, thus $M1, \langle s_1, h_2 \rangle \models K_a[a \text{ xstit}](s \wedge k)$.
- $M, \langle s_1, h_2 \rangle \models [a \text{ xstit}](s \wedge k)$ because $\langle s_1, h_2 \rangle R_a \langle s_3, h_2 \rangle$ and $M1, \langle s_3, h_2 \rangle \models s \wedge k$.

In words: from the different dynamic states based on static state s_1, through $\sim_a \circ R_{\{a\}}$, we can reach three sets of states represented by the three dotted rectangles. Semantically this means that, would h_1 be the actual history, agent a knows to be doing (knowingly sees to it that) what holds in all the dynamic states based on static state s_2 (which is $f \wedge k$), because this is the only static state which is embedded in the dotted rectangle. Would h_2 be the actual history, then agent a knows to be doing what holds in all the dynamic states based on static state s_3 (which is $s \wedge k$).

This is how $M1, \langle s_1, h_1 \rangle \models I_a[a \text{ xstit}]k$ and $M1, \langle s_1, h_2 \rangle \models I_a[a \text{ xstit}]k$ hold:

In this case the intentional equivalence class groups together two dynamic states: $\langle s_1, h_1 \rangle$ and $\langle s_1, h_2 \rangle$.

- $\langle s_1, h_1 \rangle i_a \langle s_1, h_2 \rangle$ and $M1, \langle s_1, h_2 \rangle \models [a \text{ xstit}]k$, so $M1, \langle s_1, h_2 \rangle \models I_a[a \text{ xstit}]k$.
- $\langle s_1, h_2 \rangle i_a \langle s_1, h_1 \rangle$ and $M1, \langle s_1, h_1 \rangle \models [a \text{ xstit}]k$, so $M1, \langle s_1, h_1 \rangle \models I_a[a \text{ xstit}]k$.
- The same for $\langle s_1, h_1 \rangle i_a \langle s_1, h_1 \rangle$ and $\langle s_1, h_2 \rangle i_a \langle s_1, h_2 \rangle$.
- $M, \langle s_1, h_1 \rangle \models [a \text{ xstit}]k$ because $\langle s_1, h_1 \rangle R_a \langle s_2, h_1 \rangle$ and $M1, \langle s_2, h_1 \rangle \models k$.
- $M, \langle s_1, h_2 \rangle \models [a \text{ xstit}]k$ because $\langle s_1, h_2 \rangle R_a \langle s_3, h_2 \rangle$ and $M1, \langle s_3, h_2 \rangle \models k$.

If either h_1 or h_2 would be the actual dynamic state, agent a intends to be doing what holds in all the dynamic states based on the static states s_2 and s_3 (the dynamic states in the dotted ellipse), which is k.

4 Deviant causal chains in XSTIT

4.1 Two competing theories

As mentioned in the introduction, the title 'Deviant Causal Chains in XSTIT' needs some explanation. It might be the case that it is not even possible to model deviant causal chains in XSTIT logic. That is, because the theories behind deviant causal chains and STIT logic are two different, maybe even 'competing' theories.

On the one hand there is the ontological paradigm behind the causal theory of action, which supports the view that an action is a type of event and that there is a causal link between an intention and an action. In addition, an action has certain properties such as that it takes place in time and that is has an actor. Philosophers of which can be said that they support this view are Davidson, Thomson and Goldman. From the causal theory of action and the action-as-event paradigm the deviant causal chain examples have been put forward as criticism or counterexamples as we have examined in section 2. Another important notion of this ontological view is determinism; events are deterministically caused by other events.

The difference between this theory and the theory behind STIT logic is that STIT logic elaborates on the idea that events are not deterministically caused. Furthermore, it focuses on agency, with an action considered as a property of agency as opposed to an action as an ontological object. Agency is "the relationship between an agent and the state of affairs it can bring about, without referring to it how its done, i.e. the actions performed" [40, p.1]. In STIT logic this is denoted as 'bringing it about' and 'seeing to it that'. These terms do not specifically denote *how* the agent brought about his action.

These two theories are not always going hand in hand. They are different in the way that the ontological paradigm emphasizes connections between events or actions and their consequences and the modal paradigm emphasizes connections between agents and the consequences of their choices [47]. Belnap also accentuates that these theories are competing. He stresses that Davidson played an influential role in the modal logic of agency not being popular, but that also Goldman and Thomson played a role in this [3]. He refers to this "actions as events picture" as "all ontology" and as "the dominant logical template" which "takes an agent as a wart on the skin of action, and takes an action as a kind of event" [3, p.777]. In other words, the focus lies on the action rather than the agent. Belnap even states that this view "in the case of Davidson is driven by the sort of commitment to first-order logic that counts modalities as Bad" [3, p.778]. Both quotes reassure that the ontological paradigm is not very popular within the modal paradigm and vice versa. Furthermore, in his essay Belnap tries to understand how agentive sentences can be embedded within larger contexts and he argues that the action-as-event paradigm has not contributed to his understanding and that is perhaps because "its resources do not permit it to do so" [3, p.778].

Fortunately, I am not the only one who tries to combine these theories. Xu believes that there is a union of these two theories of action which lies in "where we can directly connect actions, their agents and their consequences" [47, p.486]. He proposes a STIT theory which he argues provides such a connection. Let us see if we can find such a connection too.

4.2 The sniper example

Now that we are familiar with modeling in a XSTIT frame, this subsection will make a first attempt in modeling the sniper example. This first attempt will familiarize us with the BI-extended XSTIT frame. To recall, the sniper example is:

> "A man tries to kill somebody by shooting at him. The killer misses his victim by a mile, but the shot stampedes a herd of wild pigs that trample the intended victim to death." [16, p.78]

Which can be rewritten as:

> Agent a tries to kill agent b by shooting at agent b. Agent a misses agent b by a mile, but the shot stampedes a herd of wild pigs that trample agent b to death.

For modeling this example, I use the BI-extended XSTIT frame instead of the KI-extended XSTIT frame. In the BI-extended XSTIT frame, the knowledge operator K_a is replaced by the belief operator B_a which allows a choice to be non-successful. The result is that what an agent intentionally does, is not necessarily what happens. The semantical meaning of 'knowingly doing' is described as follows: "an agent knowingly does ϕ if ϕ holds for all the dynamic states in the epistemic equivalence set containing the *actual* dynamic state" [11, p.11]. In the case of 'believingly doing' the actual dynamic state does not need to be among the reachable epistemic equivalence set. An agent can believingly do ϕ if ϕ holds for all the dynamic states in the epistemic equivalence which does not contain the actual dynamic state. So 'believingly doing' is not closed under the "causal possibilities" an agent has [11, p.24]. This seems to fit perfectly with our sniper example.

In this example it still is the case that agent a believingly shoots agent b with the intention to kill him like in the successful sniper example in the previous section. Different to the previous frame, here I assume that agent a only has one way to kill agent b (only by shooting) and that therefore he has the choice to both believingly and intentionally shoot and kill agent b. However, the manner in which agent a had planned agent b to die is not exactly how agent b dies. I assume here that agent a neither intentionally nor believingly stampedes a herd of wild pigs, so the static state s_3 will be drawn out of the dotted rectangle and the dotted ellipse.

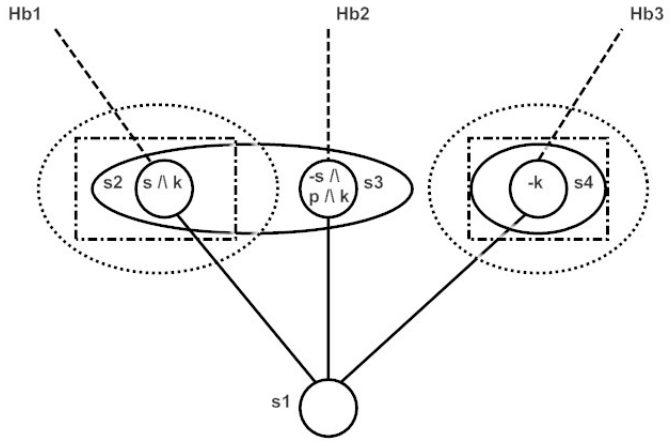

Figure 2: The sniper example in a BI-extended XSTIT model M2.

In this model it holds that $\langle s_1, h_2\rangle i_a \langle s_1, h_1\rangle$ and $\langle s_1, h_2\rangle b_a \langle s_1, h_1\rangle$. With s = 'shooting agent b', k = 'killing agent b' and p = 'stampeding a herd of wild pigs'. With 'killing agent b' it is meant that agent b's death is in some way caused by agent a (thus agent a stampeding the wild pigs and the wild pigs trampling agent b to death is considered as in some way caused by agent a). That is why in the dynamic state based on s_2 it holds that k. With 'shooting agent b' it is meant that the shot actually hits agent b (whether he dies or not is left out). In the deviant case agent a's shot did not hit agent b so that is why in the dynamic state based on s_3 it holds that $\neg s$. In the following subsection the same meanings will be used, unless otherwise specified.

At first hand, this model seems to represent the deviant causal chain example adequately. The actual history is one in the history bundle $Hb2$ and relative to this history the agent believingly and intentionally does what holds in the corresponding equivalence classes based on static state s_2. It holds that $M2, \langle s_1, h_2\rangle \models B_a[a \text{ xstit}](s \wedge k) \wedge I_a[a \text{ xstit}](s \wedge k) \wedge X(\neg s \wedge p \wedge k)$ with $h_2 \subseteq \{s_1, s_3\}$. What causally happens is not the same as the agent believingly and intentionally was seeing to. The static states s_2 and s_3 are both drawn in the same ellipse, because whether the agent eventually ends up in state s_2 or s_3 is not in his causal power. We can interpret the agent's choice as non-deterministic due to another agent which has the possibility to choose simultaneously as Broersen remarks [11]. But in this case, we can interpret the resulting static state s_3 not as resulting from the choice of another agent, but as resulting from "something that just happens, one of nature's choices" [5, p.34], or as resulting from the deviant causal chain. However, does this frame capture the essence of the sniper example? And how can we exactly describe what is happening? The next subsections try to answer these questions with help of the theorem prover SPASS.

4.3 Introduction to SPASS

SPASS is an automated full first order logic theorem prover [44].[5] In this research, SPASS can be very useful to check whether specific logical properties hold (if there can be a proof found for these) or to examine the axioms in the KI- or BI extended XSTIT frames. We can simply leave some axioms out, or we can add extra axioms, to see whether we can derive a specific logical formula. For example, we can easily check that if we only change the I_a-operator to S5 instead of KD45 in the BI-extended XSTIT frame, we will be able to derive $I_a[a \text{ xstit}]\phi \rightarrow X_a\phi$.

SPASS can take logical formulas as input, but only first order logic formulas. Because XSTIT logic is a higher order modal logic, SQEMA is used to translate the higher order modal formulas into their first-order modal correspondences.[6] SQEMA leaves one variable unquantified in the first-order formula. This variable will be bound by a universal quantifier.

For representing the KI- and BI-extended XSTIT frames, I introduce five modalities in SPASS, which are corresponding to the operators I, B or K, H (for historical necessity or possibility, represented by $\Box$ and $\Diamond$ in [11]), XSTIT (seeing to it that) and X (next state relation denoted by R_X in [11]). These modalities are each corresponding to a number and denoted respectively as [0], $\langle 0\rangle$, [1], $\langle 1\rangle$, etc., for their box or diamond representation.

In the examples we will only consider one agent, so axioms considering multiple agents are left out. This results in the following axioms and logical properties being used: [7]

[5] I used SPASS version 3.5, which can be downloaded from http://www.spass-prover.org/.

[6] SQEMA can be used as web-interface at www.fmi.uni-sofia.bg/fmi/logic/sqema/.

[7] The axioms (p), (Lin), (Sett) and (XSett) are from [13, p.470] where the axiom schemes are composed for a

KI-extended XSTIT frame:	BI-extended XSTIT frame:
(p): $p \rightarrow \Box p$ for every proposition	(p): $p \rightarrow \Box p$ for every proposition
(Lin): $\neg X \neg \phi \leftrightarrow X\phi$	(Lin): $\neg X \neg \phi \leftrightarrow X\phi$
(Sett): $\Box X\phi \rightarrow [a \text{ xstit}]\phi$	(Sett): $\Box X\phi \rightarrow [a \text{ xstit}]\phi$
(XSett): $[a \text{ xstit}]\phi \rightarrow X\Box\phi$	(XSett): $[a \text{ xstit}]\phi \rightarrow X\Box\phi$
(KX): $K_a X\phi \rightarrow K_a[a \text{ xstit}]\phi$	
(ER): $K_a[a \text{ xstit}]\phi \rightarrow XK_a\phi$	(B-ER): $B_a[a \text{ xstit}]\phi \rightarrow XB_a\phi$
(Unif-str): $\Diamond K_a[a \text{ xstit}]\phi \rightarrow K_a\Diamond[a \text{ xstit}]\phi$	(B-Unif-str): $\Diamond B_a[a \text{ xstit}]\phi \rightarrow B_a\Diamond[a \text{ xstit}]\phi$
(K-S): $\Box K_a\phi \leftrightarrow K_a\Box\phi$	(B-S): $\Box B_a\phi \rightarrow B_a\Box\phi$
(X-Eff-I): $\Box K_a X\phi \rightarrow I_a[a \text{ xstit}]\phi$	(BX-Eff-I): $\Box B_a X\phi \rightarrow I_a[a \text{ xstit}]\phi$
(I $\Rightarrow$ K): $I_a[a \text{ xstit}]\phi \rightarrow K_a[a \text{ xstit}]\phi$	(I $\Rightarrow$ B): $I_a[a \text{ xstit}]\phi \rightarrow B_a[a \text{ xstit}]\phi$
I_a: S5	I_a: KD45
K_a: S5	B_a: KD45
Historical necessity: S5	Historical necessity: S5
$p \rightarrow \Box p$ for every atomic proposition p	$p \rightarrow \Box p$ for every atomic proposition p

The axioms (Agg) $[a \text{ xstit}]\phi \wedge [a \text{ xstit}]\psi \rightarrow [a \text{ xstit}](\phi \wedge \psi)$ and (Mon) $[a \text{ xstit}](\phi \wedge \chi) \rightarrow [a \text{ xstit}]\phi$ can be derived from the axioms above so they do not need to be added explicitly. Also the fact that the XSTIT- and X-relations are in KD follows from the axioms above. See the attachments for the basic SPASS-files[8] of the KI-extended and the BI-extended XSTIT frames with an example conjecture for which a proof can be found.[9]

Every modal formula specified in the axiom-part of the SPASS-file is translated into [u]formula, where [u] is the universal modality. That is why local assumptions which only hold in the dynamic state we are evaluating from, are given in the conjecture-part. For example if $X(s)$ will be part of *list_of_special_formulae(axioms,em1)* then a proof can be found for $B_a[a \text{ xstit}]s$ and $I_a[a \text{ xstit}]s$ (thus only $B_a[a \text{ xstit}]s$ or $I_a[a \text{ xstit}]s$ as conjecture). That is because $X(s)$ will be bound by a universal modality. While if we imagine to be in a specific dynamic state, if next happens s, in both the KI- and BI-extended XSTIT frame it is not always the case that then also $B_a[a \text{ xstit}]s$ or $I_a[a \text{ xstit}]s$ hold. If we leave out $X(s)$ as part of the axioms and make it part of the conjecture, thus try to proof $X(s) \rightarrow B_a[a \text{ xstit}]s$ or $X(s) \rightarrow I_a[a \text{ xstit}]s$, no proof can be found for any of these formulas and that is what we want.

Note that SPASS may return 'completion found' if it is discovered that a proof cannot be found, but it can also run forever if a proof cannot be found. Since the formulas I want to proof are not very complicated, I do not expect SPASS to run for more than fifteen minutes. Of course the time SPASS is running depends on the computer, but this time indication is just to give an idea. Most of the times if a proof was found it took less than one minute.

4.4 Possible representations

First of all, important to note is that an intentional action is not always successful when a deviant causal chain is involved. The deviant causal chain might instantiate another event, which does not have the same effect. Thus $I_a[a \text{ xstit}]\phi \rightarrow X\phi$ should not hold. And that is why the BI-extended XSTIT frame can be a perfect candidate for modeling deviant causal chain examples. Because the I_a-operator and the B_a-operator are both in KD45 and because we have weaker axioms for

single agent. The axioms for the K_a-operator are from [11, p.12] and for the B_a-operator from [11, p.24]

[8] SPASS-files have extension '.dfg'.

[9] For explanation of the syntax of SPASS I refer to [42].

the B_a-operator, we cannot derive that what an agent intentionally does is necessarily what happens. The logical properties of our operators are thus very important for modeling deviant causal chains. In general, it is very important to take into account what certain logical properties are implying for the world we are modeling in. This is an example of what logical properties are implying for our real life situation. Both the I_a-operator and the B_a-operator should be in KD45. If one of them would be in S5 we could make the following derivations to proof $I_a[a \text{ xstit}]\phi \rightarrow X\phi$:

I in KD45, B in S5	I in S5, B in KD45:
$I_a[a \text{ xstit}]\phi \rightarrow B_a[a \text{ xstit}]\phi$ (I $\Rightarrow$ B)	
$B_a[a \text{ xstit}]\phi \rightarrow [a \text{ xstit}]\phi$ (S5, reflexivity)	$I_a[a \text{ xstit}]\phi \rightarrow [a \text{ xstit}]\phi$ (S5, reflexivity)
$[a \text{ xstit}]\phi \rightarrow X\Box\phi$ (XSett)	$[a \text{ xstit}]\phi \rightarrow X\Box\phi$ (XSett)
$X\Box\phi \rightarrow X\phi$ (standard modal reasoning)	$X\Box\phi \rightarrow X\phi$ (standard modal reasoning)

Another property, which we would ascribe to our intentional actions, can be found in our logical framework. That is, a proof can be found for the following: $I_a[a \text{ xstit}]p \wedge I_a[a \text{ xstit}](p \rightarrow k) \rightarrow I_a[a \text{ xstit}]k$. If we are intentionally doing ϕ and we are also intentionally seeing to it that if ϕ holds then also ψ holds, then we are also intentionally doing ψ.

Now, let us first investigate an example where the deviant causal chain takes place, but where the deviant causal chain does not bring about the same event as the agent wanted to bring about. It will be easier to model this example and from here on we might extend or adapt the frame to model the sniper example. The example frame F3 models the following example:

> Agent a tries to kill agent b by shooting at him. Agent a misses his victim by a mile and the shot stampedes a herd of wild pigs that run away, so that agent b is still alive.

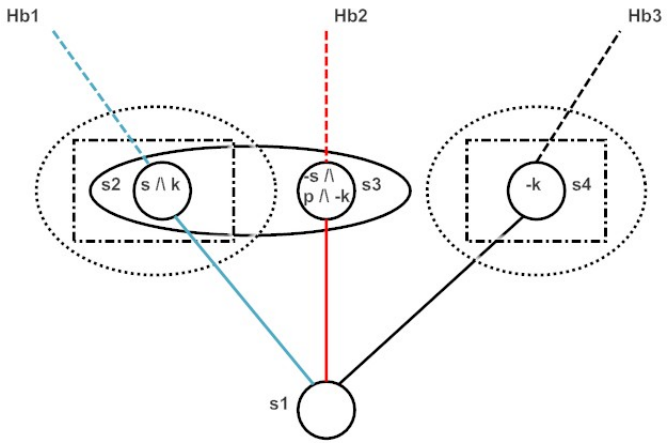

Figure 3: Not killing in BI-extended XSTIT model M3.

From now on, the history bundle which the actual history is running through is represented by red lines. This history $h_2 \in Hb2$ where $h_2 \subseteq \{s_1, s_3\}$ contains the actual dynamic state where we assume the agent is ending up in, which is dynamic state $\langle s_3, h_2 \rangle$. The history bundle which contains the dynamic state where the agent believes to end up is represented by blue lines. This is dynamic state $\langle s_2, h_1 \rangle$ with $h_1 \in Hb1$ where $h_1 \subseteq \{s_1, s_2\}$.

Relative to history h_2 the agent believingly and intentionally exercise the choice represented by respectively the dotted rectangle and the dotted ellipse around static state s_2. Thus it holds that $M3, \langle s_1, h_2 \rangle \models B_a[a \text{ xstit}](s \wedge k) \wedge I_a[a \text{ xstit}](s \wedge k)$. Due to the deviant causal chain we end up in a dynamic state which is not among the epistemic accessible states. It holds that $M3, \langle s_1, h_2 \rangle \models X(\neg s \wedge p \wedge \neg k)$.

As Broersen remarks the question of whether or not the agent killed intentionally does not depend on the outcome of the action [11, p.25]. Although agent a's shot actually missed agent b, the fact that agent b is not dead does not change the fact that agent a actually performed the intentional act of killing agent b. We can perfectly state that all of the following holds: $M3, \langle s_1, h_2 \rangle \models B_a[a \text{ xstit}](s \wedge k) \wedge I_a[a \text{ xstit}](s \wedge k) \wedge X(\neg s \wedge p \wedge \neg k)$. Thus, we can model a situation where a deviant causal chain does not have the same effect as the agent was originally intending. That is, because we obtain no contradictory logical formulas. While the agent performed the action of intentionally killing his victim, in the end this does not matter since agent b was not killed or hurt at all. But it does matter if agent b would be hurt indeed, such as in the sniper example. Now let us investigate our previous frame F2 again.

The problem with the deviant causal chain example, in this case the sniper example, is that in the philosophical literature it is assumed that the sniper, in the end, did not intentionally kill his victim. Thus although he initially intentionally was seeing to it that k, after the deviant causal chain took place, we conclude that he did not intentionally see to it that k, while still k took place. Thus we obtain $I_a[a \text{ xstit}]k$, $X(\neg s \wedge p \wedge k)$ and $\neg I_a[a \text{ xstit}]k$ which can logically not hold together.

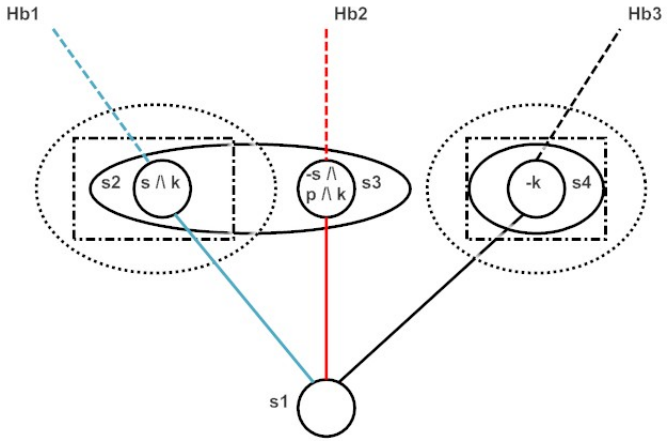

Figure 4: The sniper example in BI-extended XSTIT model M2 with color.

In this model, it is logically not possible that $M2, \langle s_1, h_2 \rangle \models B_a[a \text{ xstit}](s \wedge k) \wedge$
$I_a[a \text{ xstit}](s \wedge k) \wedge X(\neg s \wedge p \wedge k) \wedge \neg I_a[a \text{ xstit}](s \wedge k)$ hold with $h_2 \in Hb_2$ where $h_2 \subseteq \{s_1, s_3\}$. However, it is possible to state that the following holds: $M2, \langle s_1, h_2 \rangle \models B_a[a \text{ xstit}](s \wedge k) \wedge$
$I_a[a \text{ xstit}](s \wedge k) \wedge X(\neg s \wedge p \wedge k)$. So we are actually able to model the sniper example in this logic, but then we have to conclude that in this example the sniper did *intentionally* kill his victim, which is contradicting with the philosophical literature.

It might be a solution to state that the agent indeed did 'personally kill' his victim, but in the

end this did not happen, it was the case that the victim was 'killed by wild pigs', denoted by a different atomic proposition. Thus the action the agent wanted to perform, did not eventually take place. However, this does not change the fact that in both static states the same atomic proposition for 'agent b is dead' is true and that the agent was intentionally seeing to it that 'agent b is dead', but that after the deviant causal chain took place, the agent did not. So we obtain the same contradiction.

Although the BI-extended XSTIT frame seems suitable, our previous frames could not model the deviant causal chain example where the agent's intended action takes place, while we have to conclude that the agent did not act intentionally. However, we can easily model an agent believingly and intentionally *using* the deviant causal chain as follows:

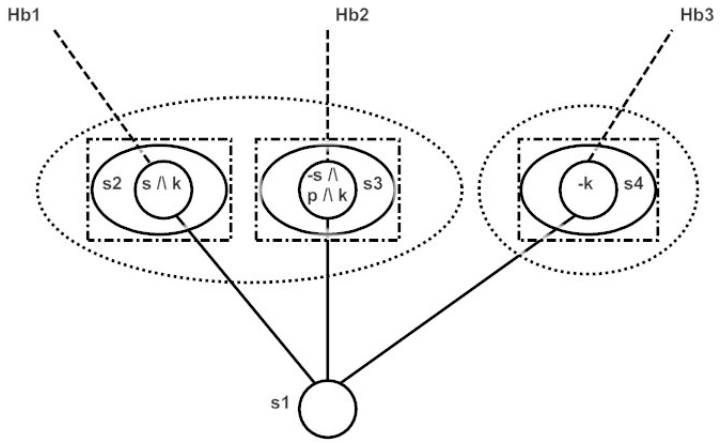

Figure 5: Using the deviant causal chain in BI-extended XSTIT model M4.

Here we can notice again the practical implications of the difference between believingly and intentionally doing. Agent a can choose to intentionally kill by believingly performing two different actions. Assume he uses the deviant causal chain, the actual dynamic state will be $\langle s_3, h_2 \rangle$ and it holds that $M4\langle s_1, h_2 \rangle \models I_a[a \text{ xstit}]k \wedge B_a[a \text{ xstit}](\neg s \wedge p \wedge k) \wedge X(\neg s \wedge p \wedge k)$ with $h_2 \in Hb_2$ where $h_2 \subseteq \{s_1, s_3\}$. In this case agent a kills agent b by stampeding the herd of wild pigs, but he could also choose to kill agent b by shooting him.

In model M4 we assume that the deviant causal chain will always succeed in killing agent b. It is, however, also possible that the wild pigs do not kill agent b. This we can model as follows:

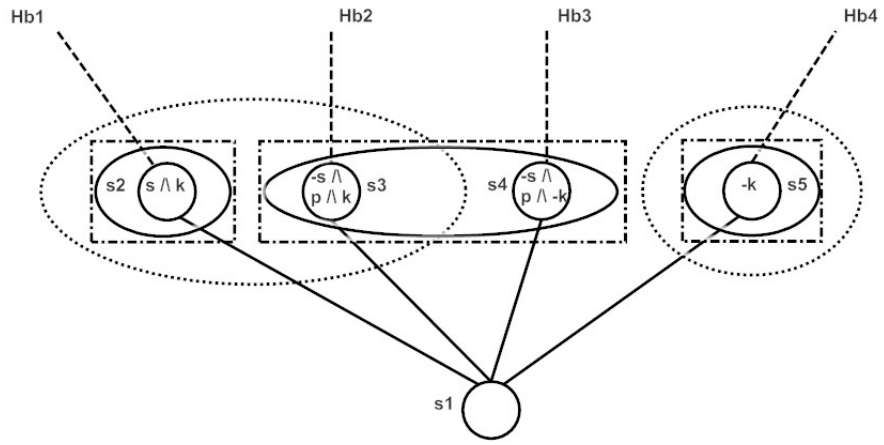

Figure 6: Using the deviant causal chain with possible failure in BI-extended XSTIT model M5.

Of course we can think of more frames which can model this situation, but this frame gives us a first impression. With this frame we would have to drop the axiom (I $\Rightarrow$ B) because the dotted rectangle is not contained in any dotted ellipse. Anyhow, it is good to take into account the question if we really need this axiom, because we can think of situations where we can believingly act without intentionally seeing to it that, such as in this example. In this case, our agent can believingly stampede the herd of wild pigs in two different ways, but he himself cannot influence if the wild pigs will trample agent b to death (because both static state s_3 and s_4 are contained in the same ellipse). Agent a can choose to stampede the herd of wild pigs with the intention to kill him. He can also choose to shoot the pigs in a way that they likely do not run over agent b, but then he will not do that with any intention. If the wild pigs will eventually kill agent b is not in agent a's causal power. We can think of the situation where agent a stampedes the herd of wild pigs with the intention to kill him, but the wild pigs will run the other way as to avoid agent b, thus agent b will be still alive. The following holds in this situation with the actual dynamic state $\langle s_4, h_3\rangle$: $M5\langle s_1, h_3\rangle \models I_a[a \text{ xstit}]k \wedge B_a[a \text{ xstit}]p \wedge X(\neg s \wedge p \wedge \neg k)$ with $h_3 \in Hb3$ where $h_3 \subseteq \{s_1, s_4\}$.

We can model other deviant causal chain examples, but still we have not modeled the one where the agent's intended action did take place in a non-intended way. At least, we cannot model it when the action takes place within one 'step'. Maybe it is useful however, to examine a frame which models the sniper example in multiple steps. Therefore, consider the following frame:

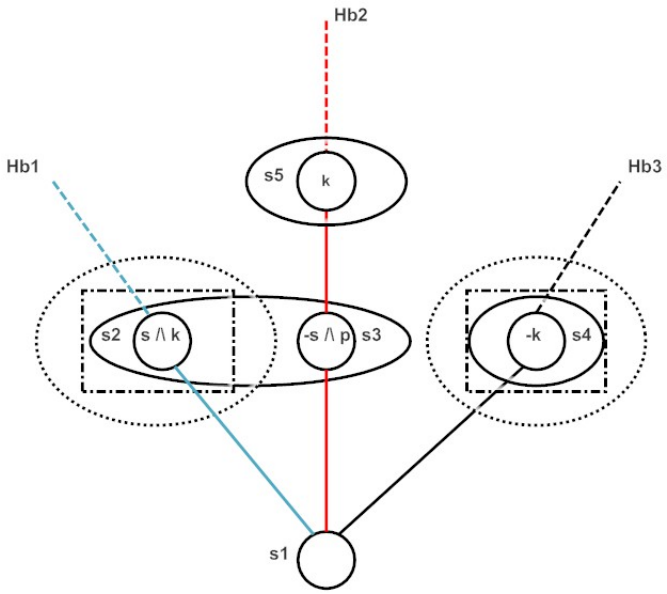

Figure 7: The sniper example in multiple steps in BI-extended XSTIT model M6.

The first step is the same as before where the agent is believing to end up in dynamic state $\langle s_2, h_1\rangle$ with $h_1 \in Hb_1$ where $h_1 \subseteq \{s_1, s_2\}$. But then the deviant causal chain takes places. It holds that $M6, \langle s_1, h_2\rangle \models B_a[a \text{ xstit}](s \wedge k) \wedge I_a[a \text{ xstit}](s \wedge k) \wedge X(\neg s \wedge p)$ with $h_2 \in Hb_2$ where $h_2 \subseteq \{s_1, s_3, s_5\}$. It means that the killing, at least in one step, did not take place, although the agent was intentionally seeing to that. In static state s_3 the agent does not have a choice but to go to static state s_5: $M6, \langle s_3, h_2\rangle \models \Box X(k)$. That is because we assume that the agent cannot influence the deviant causal chain, he cannot influence the way the wild pigs trample the victim to death. But it also holds that $M6, \langle s_3, h_2\rangle \models \neg I_a[a \text{ xstit}]k$, so the agent did not intentionally kill, although in his first step (from $\langle s_1, h_2\rangle$ to $\langle s_3, h_2\rangle$) he intentionally was seeing to it. So we would describe the sniper example as follows: $M6, \langle s_1, h_2\rangle \models B_a[a \text{ xstit}](s \wedge k) \wedge$
$I_a[a \text{ xstit}](s \wedge k) \wedge X(\neg s \wedge p) \wedge X(X(k)) \wedge X(\neg I_a[a \text{ xstit}]k)$.

In my opinion, this last logical formula does not really present the deviant causal chain adequately. That is, because we cannot derive from the frame or formula that the two actions are related to each other, so how should we know it presents a deviant causal chain? The other downside of modeling the sniper example in multiple steps is that suddenly we can consider many more possible choices. We can imagine the situation where the wild pigs do not trample the victim to death. We can also think about ways in how the agent can actually influence the deviant causal chain. For example, we can imagine that the sniper, after he shot, may react and scream 'Watch out!', with as result that his victim jumps away and that he is not killed by the wild pigs. Modeling causal deviancy in multiple steps has some positive sides too. It comes closer to real life situations, where it is also the case that many deviant causal chains can possibly happen. Furthermore, if we examine multiple steps, we can more clearly see the 'chain' of events which takes place and therefore simulate causal chains in XSTIT logic better. Consider the following frame for a possible representation:

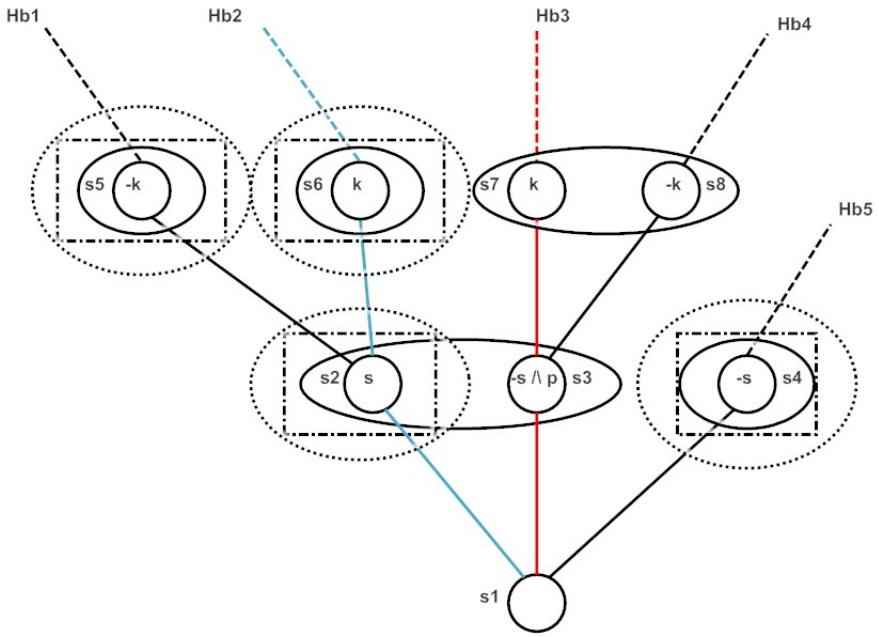

Figure 8: The sniper example in multiple steps in BI-extended XSTIT model M7.

In this frame, the agent cannot believingly choose to use the deviant causal chain and it is possible that the pigs either kill or not kill agent b, but that is not in agent a's causal power. Our agent has the choice to intentionally shoot and has, after this action took place, the choice to somehow not kill agent b (by warning him for example).

In all frames proposed, the problem remains that we cannot really derive from the frame or logical formulas that we are dealing with causal deviancy, as it might well be the case that another agent is choosing simultaneously in the situations where more static states are drawn within one ellipse. In the frames where the agent is using the deviant causal chain, we cannot make out from looking at the frame that a causal deviancy is being modeled.

4.5 Is it possible?

The BI-extended XSTIT frame is suitable for modeling certain cases of causal deviancy. It can model causal deviancy where the causal chain leads to a different outcome than the agent intended, or where the agent did intentionally make use of the deviant causal chain to obtain his goal. However, the BI-extended XSTIT frame cannot model the famous causal deviancy examples from the philosophical literature, such as the sniper example. We obtain a contradiction if we assume that the agent performs the action of intentionally ϕ-ing, but due to the deviant causal chain we classify his action as not intentional. If we model the sniper example in our BI-extended XSTIT frame, we have to conclude that the agent actually did ϕ *intentionally.*

Modeling the sniper example in multiple steps is not the best solution, because we would have to examine at least two 'levels' of the frame to describe the causal deviant case. In addition, it seems that with modeling in multiple steps, the central issue of the causal deviant cases is lost: the issue of performing a certain act intentionally but yet not have performed that act intention-

ally *at the same time*. Furthermore, in this way of modeling, we suddenly have to consider many more possible actions.

I conclude that modeling the causal deviancy examples is not possible in the BI-extended XSTIT frame, but further research can examine how it might be possible. I suggest the next step would be to investigate if deviant causal chain examples can be modeled in probabilistic XSTIT logic, XSTIT^p logic [12, 13, 14]. The XSTIT^p framework might be very promising for this research because it allows a choice to be unsuccessful in an elegant way. It takes the agent's beliefs about the result of his choice and expresses this belief in probabilities. Thus it combines STIT logic with "probabilities in the object language, enabling us to say that an agent exercises a choice for which it believes to have a chance higher than c to see to it that ϕ results in the next state" [13, p.522]. We can imagine that in the sniper example the agent believes he has a high change of killing agent b by shooting at him.

A relatively big time-consuming part of this research turned out to be the examination of theorem prover SPASS and understanding how XSTIT-logic can be represented in SPASS. In the end, it appeared that I did not have to use SPASS as much as I expected to, partly because of the conclusion that it is not possible to model deviant causal chain examples in XSTIT logic. However, this research has provided a basic template and a good example of how SPASS can be used. The SPASS files in the appendices might be of use in further research in modeling deviant causal chain examples.

5 Conclusion

This thesis has investigated the place of deviant causal chain examples within the philosophy of action. It has provided us with definitions and examples. Furthermore, we have been informed about the arguments given in the debate about whether or not causal deviancy is a real problem for the causal theory of action. Next, we have modeled examples in KI- and BI-extended XSTIT frames and we have investigated how these logical frameworks can be used. Finally, this thesis has shed light on how a theorem prover can be useful in this kind of research and we have seen different possible representations of deviant causal chains in XSTIT logic.

The causal theory of action is the philosophical theory which assumes that actions are events and that an event is an action if and only if it is caused and rationalized by the agent's intention. Basic and non-basic causal deviant situations are situations where a control undermining event takes place between respectively the agent's intention and the basic action, or the basic action and the non-basic action. Deviant causal chain examples were first proposed by Chisholm and brought to further discussion by Davidson. In philosophical literature, causal deviant examples are mostly examples where the event which takes place is the same as the agent in question initially intended. However, due to the deviant causal chain happening of which the agent did not know beforehand, in the examples proposed we have to conclude that the agent did not perform that particular act intentionally. These examples demonstrate that the causation and rationalization of the agent's intention is not a sufficient condition for an event being an action, or for an action being intentional.

Several solutions have been proposed to save the causal theory of action from causal deviancy. Solutions examined in this thesis are the following, where in order for an action to be intentional the action should...

- happen according to the agent's plan (events involved, or the time of happening);
- consist of intentional undertaken actions, if the action consists of more steps being taken;
- guided by the contents of mental attitudes;
- be an action in which the intention persist into the time of action;
- be proximately or directly caused by an action (causal immediacy strategy);
- be a response to the reason for which the action is executed for (sensitivity strategy);
- be responsive to the content of the mental state;
- take place according to the intentional deviant causal chains specified by neurosciences;

Of these investigated solutions, the sensitivity strategy seems the best solution thus far. However, arguments have been proposed against this strategy too, which reminds us that we cannot take any solution for granted. The question remains which solution is really the best and this question cannot be answered yet. Nevertheless, I conclude that the causal theory of action should not be abandoned despite of the difficulties the deviant causal chain has brought forward, because many philosophers have argued that their solutions save the causal theory of action, or that causal deviancy is no problem any more. Although we might not yet have established a causal theory of action which can deal with causal deviancy, it seems we are heading to one with all

the improvements made thus far.

Modeling deviant causal chains in XSTIT logic is challenging because the theory behind deviant causal chains emerges from an ontological paradigm, while the theory behind XSTIT logic emerges from a modal paradigm. XSTIT logic is a variation on STIT logic, which is a philosophical logic of agency. The KI- and BI-extended XSTIT frame have been used in this research together with theorem prover SPASS to investigate possible representation of the sniper example in XSTIT logic. SPASS appeared to be very suitable to examine derivations in this logic.

It it not possible to model one of the deviant causal chain examples, the sniper example, in the BI-extended XSTIT frame in one step. The method to model causal deviancy in multiple steps does not really capture the essence of the deviant causal chain examples and examining multiple steps gives us the impression that the agent has a choice after the deviant causal chain took place. In the philosophical literature it seems that the agent cannot influence the deviant causal chain. Moreover, if we include multiple steps, many more possible actions are involved, which can make the model confusing.

Further research is necessary in modeling deviant causal chains in XSTIT logic. I suggest to investigate if the philosophical examples of causal deviancy can be modeled in XSTIT^p logic and to consider modeling in multiple steps in this logic as well. It would also be interesting to investigate modeling causal deviancy in other forms of STIT logic if only to test the properties of the logical frameworks.

Some causal deviant examples were not examined by this research. Further research can investigate if basic deviancy can be modeled in STIT logic, or deviancy where the causal chain runs through another person. Moreover, many other interesting examples are provided by philosophical literature such as the side-effect problem. It would be interesting to test whether such an example can be adequately modeled.

The research question if causal relations can be modeled in STIT logic, has to be left unanswered. This research did just examine one example within one logical STIT framework and therefore cannot answer this question. However, I conclude that the possibility does not have to be rejected, since no main problems have arisen so far. I am still hopeful that it is possible to obtain a union between the philosophical causal theory of action and STIT logic.

In this thesis, two main sub domains of Artificial Intelligence have been studied: philosophy and logic. Firstly, this research has examined the philosophical debate of causal deviancy in the context of the causal theory of action. Secondly, it has demonstrated the possibilities of STIT logic in terms of modeling agents and agency in real life situations. Although I cannot conclude that a union between the causal theory of action and STIT logic is obtained, at least a union between these two main sub domains has been established. The conclusion that the modeling of deviant causal chains in XSTIT logic does not have to be rejected, reveals that the study of Artificial Intelligence is able to reunite different theories. Hopefully this thesis will stimulate us to continue studying problems and questions in Artificial Intelligence from different perspectives.

References

[1] Anscombe, G. *Intention*. Cornell University Press, 1957.

[2] Beck, L. *The actor and the Spectator*. Yale University Press, London, 1975.

[3] Belnap, N. Backwards and forwards in the modal logic of agency. *Philosophy and Phenomenological Research 51*, 4 (1991), 777–807.

[4] Belnap, N., and Perloff, M. Seeing to it that: A canonical form for agentives. *Theoria 54*, 3 (1988), 175–199.

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

[6] Bishop, J. Peacocke on intentional action. *Analysis 41*, 2 (1981), 92–98.

[7] Bishop, J. *Natural Agency: An Essay on the Causal Theory of Action*. Cambridge University Press, Cambridge, 1989.

[8] Brand, M. *Intending and Acting: Toward a Naturalized Action Theory*. Cambridge University Press, Cambridge, 1984.

[9] Bratman, M. *Intention, Plans, and Practical Reason*. The David Hume Series: Philosophy and Cognitive Science Reissues. Center for the Study of Language and Information Publication, 1987.

[10] Broersen, J. First steps in the stit logic analysis of intentional action. In *Proceedings ESSLLI 2009 workshop on Logical Methods for Social Concepts (LMSC'09)* (2009), A. Herzig and E. Lorini, Eds. informal proceedings.

[11] Broersen, J. Making a start with the stit logic analysis of intentional action. *Journal of philosophical logic 40*, 4 (August 2011), 499–531.

[12] Broersen, J. Modeling attempt and action failure in probabilistic stit logic. In *Proceedings of Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011)* (2011), T. Walsh, Ed., IJCAI, pp. 792–797.

[13] Broersen, J. Probabilistic stit logic. In *Proceedings 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2011)* (2011), W. Liu, Ed., vol. 6717 of *Lecture Notes in Artificial Intelligence*, Springer, pp. 521–531.

[14] Broersen, J. Probabilistic stit logic and its decomposition. *International Journal of Approximate Reasoning 54* (2013), 467477.

[15] Chisholm, R. Freedom and action. In *Freedom and Determinism*, K. e. Lehrer, Ed. Random House, Oxford, 1966, pp. 28–44.

[16] Davidson, D. *Essays on Actions and Events*. Clarendon Press, Oxford, 1980.

[17] Ducasse, C. Explanation, mechanism, and teleology. *Journal of Philosophy 22*, 6 (1925), 150–155.

[18] Enç, B. Causal theories and unusual causal pathways. *Philosophical Studies 55*, 3 (1989), 231–261.

[19] Frankfurt, H. The problem of action. *American Philosophical Quarterly 15* (1978), 157–162.

[20] Ginet, C. On action. *Cambridge University Press* (1990).

[21] Goldman, A. I. *A Theory of Human Action*. Englewood Cliffs, 1970.

[22] Keil, G. What do deviant causal chains deviate from? In *Intentionality, Deliberation and Autonomy: the action-theoretic basis of practical philosophy*, C. Lumer and S. Nannini, Eds. Aldershot (Ashgate), 2007.

[23] Lowe, E. *A Survey of Metaphysics*. Oxford University Press, 2002.

[24] Mayr, E. Deviant causal chains. In *Understanding Human Agency* (2011), Oxford University Press, pp. 104–141.

[25] Mele, A. Intentional action and wayward causal chains: the problem of tertiary waywardness. *Philosophical Studies* (1987), 55–60.

[26] Mele, A. *Springs of Action*. Oxford University Press, New York, 1992.

[27] Mele, A. *Motivation and Agency*. Oxford University Press, Oxford, 2003.

[28] Mitchell, D. Deviant causal chains. *American Philosophical Quarterly 19*, 4 (October 1982), 351–353.

[29] Montmarquet, J. A. Causal deviancy and multiple intentions. *Analysis 42*, 2 (1982), 106–110.

[30] Morton, A. Because he thought he had insulted him. *Journal of Philosophy 72* (1975), 5–15.

[31] Peacoke, C. *Holistic explanation: action, space, interpretation*. Oxford University Press on Demand, 1979.

[32] Prior, A. *Past, Present and Future*. Oxford University Press, Oxford, 1967.

[33] Schlosser, M. Basic deviance reconsidered. *Analysis 67*, 3 (2007), 186–194.

[34] Searle, J. *Intentionality*. Cambridge University Press, Cambridge, 1983.

[35] Shaffer, J. *Philosophy of mind*. Foundations of philosophy series. Prentice-Hall, 1968.

[36] Stout, R. Deviant causal chains. In *A Companion to the Philosophy of Action*, T. O'Connor and C. Sandis, Eds., Blackwell Companions to Philosophy. Wiley, 2012.

[37] Tännsjö, T. On deviant causal chains - no need for a general criterion. *Analysis 68*, 3 (2009), 469–473.

[38] Taylor, C. *The Explanation of Behaviour*. International library of philosophy and scientific method. Routledge & Kegan Paul, 1964.

[39] Thalberg, I. Do our intentions cause our intentional actions? *American Philosophical Quarterly 21*, 3 (1984), 249–260.

[40] Troquard, N., Trypuz, R., and Vieu, L. Towards an ontology of agency and action from stit to ontostit+. In *Proceedings of the 2006 conference on Formal Ontology in Information Systems: Proceedings of the Fourth International Conference (FOIS 2006)* (Amsterdam, The Netherlands, 2006), IOS Press, pp. 179–190.

[41] van Ditmarsch, H., van der Hoek, W., and Kooi, B. Epistemic logic. In *Dynamic Epistemic Logic*, vol. 337. Springer, 2008, pp. 11–42.

[42] Weidenbach, C. Spass input syntax version 1.5.

[43] Weidenbach, C., Dimova, D., Fietzke, A., Kumar, R., Suda, M., and Wischnewski, P. Spass version 3.5. In *Automated Deduction CADE-22*, R. Schmidt, Ed., vol. 5663 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2009, pp. 140–145.

[44] Weidenbach, C., Schmidt, R. A., Hillenbrand, T., Rusev, R., and Topic, D. System description: Spass version 3.0. In *Proceedings of the 21st international conference on Automated Deduction: Automated Deduction* (Berlin, Heidelberg, 2007), CADE-21, Springer-Verlag, pp. 514–520.

[45] Wilson, G. The intentionality of human action. *Stanford University Press* (1989).

[46] Wittgenstein, L., Hacker, P., and Schulte, J. *Philosophical Investigations*. Wiley, 2010.

[47] Xu, M. Combinations of *Stit* and actions. *Journal of Logic, Language and Information 19*, 4 (2010), 485–503.

Appendices

A BI-frame

SPASS-file for standard BI-frame with to proof $I_a[a \text{ xstit}]s \wedge I_a[a \text{ xstit}](s \to k) \to I_a[a \text{ xstit}]k$

```
begin_problem(BI_frame).
list_of_descriptions.
  name({* BI_frame *}).
  author({* Nadine Hermans *}).
  status(unknown).
  description({*

        To proof: I[XSTIT]s /\ I[XSTIT](s->k) -> I[XSTIT]k

  *}).
end_of_list.

list_of_symbols.
predicates[(s,0), (k,0),
          (I,0), (B,0), (H,0), (XSTIT,0), (X,0),
          (R0,2), (R1,2), (R2,2), (R3,2), (R4,2)].
translpairs[(I,R0), (B,R1), (H,R2), (XSTIT,R3), (X,R4)].
end_of_list.

list_of_special_formulae(axioms, eml).

% Lin: -X-p <-> Xp
% <4>p <-> [4]p
% forall y1((-(xR4y1) \/ forall z1((xR4z1) -> (z1 = y1))) /\ exists z2(xR4z2))
formula(
       forall([y1,x],
              and(
                     or(not(R4(x,y1)),
                            forall([z1],
                                   implies(R4(x,z1),equal(z1,y1))
                            )
                     ),
                     exists([z2], R4(x,z2))
              )
       ), Lin
).

% Sett: []Xp ->[XSTIT]p
% [2][4]p -> [3]p
% forall z1((xR3z1) -> exists z2((xR2z2) /\ (z2R4z1)))
formula(
       forall([z1,x],
              implies(R3(x,z1),
                     exists([z2],
                            and(R2(x,z2),R4(z2,z1))
                     )
```

```
47                )
48        ), Sett
49 ).
50
51 % XSett: [XSTIT]p -> X[]p
52 % [3]p -> [4][2]p
53 % forall z1((xR4z1) -> forall z2((z1R2z2) -> (xR3z2)))
54 formula(
55        forall([z1,x],
56                implies(R4(x,z1),
57                        forall([z2],
58                                implies(R2(z1,z2),R3(x,z2))
59                        )
60                )
61        ), XSett
62 ).
63
64 % B-ER: B[XSTIT]p -> XBp
65 % [1][3]p -> [4][1]p
66 % forall z1((xR4z1) -> forall z2((z1R1z2) -> exists z3((xR1z3) /\ (z3R3z2))))
67 formula(
68        forall([z1,x],
69                implies(R4(x,z1),
70                        forall([z2],implies(R1(z1,z2),
71                                exists([z3],and(R1(x,z3),R3(z3,z2)))
72                                )
73                        )
74                )
75        ), B_ER
76 ).
77
78 % B-Unif-Str: <>B[XSTIT]p -> B<>[XSTIT]p
79 % <2>[1][3]p -> [1]<2>[3]p
80 % forall y1(-(xR2y1) \/ forall z1((xR1z1) -> exists z2((z1R2z2) /\ forall z3((z2R3z3)
       -> exists z4((y1R1z4) /\ (z4R3z3))))))
81 formula(
82        forall([y1,x],
83                or(not(R2(x,y1)),
84                        forall([z1],
85                                implies(R1(x,z1),
86                                        exists([z2],
87                                                and(R2(z1,z2),
88                                                        forall([z3],
89                                                                implies(R3(z2,z3),
90                                                                        exists([z4],
91                                                                                and(R1(y1,z4),R3(z4,z3))
92                                                                        )
93                                                                )
94                                                        )
95                                                )
96                                        )
97                                )
98                        )
```

```
99              )
100       ), B_Unif_Str
101 ).
102 
103 % B-S: []Bp -> B[]p
104 % [2][1]p -> [1][2]p
105 % forall z1((xR1z1) -> forall z2((z1R2z2) -> exists z3((z3R1z2) /\ (xR2z3))))
106 formula(
107       forall([z1,x],
108              implies(R1(x,z1),
109                     forall([z2],
110                            implies(R2(z1,z2),
111                                   exists([z3],
112                                          and(R1(z3,z2),R2(x,z3))
113                                   )
114                            )
115                     )
116              )
117       ), B_S
118 ).
119 
120 % BX-Eff-I: []BXp -> I[XSTIT]p
121 % [2][1][4]p -> [0][3]p
122 % forall z1((xRz1) -> forall z2((z1R3z2) -> exists z3((z3R4z2) /\ exists z4((z4R1z3) /\
    (xR2z4)))))
123 formula(
124       forall([z1,x],
125              implies(R0(x,z1),
126                     forall([z2],
127                            implies(R3(z1,z2),
128                                   exists([z3],
129                                          and(R4(z3,z2),
130                                                 exists([z4],
131                                                        and(R1(z4,z3),R2(x,z4))
132                                                 )
133                                          )
134                                   )
135                            )
136                     )
137              )
138       ), BX_Eff_I
139 ).
140 
141 % (I => B): I[XSTIT]p -> B[XSTIT]p
142 % [0][3]p -> [1][3]p
143 % forall z1((xR1z1) -> forall z2((z1R3z2) -> exists z3((xRz3) /\ (z3R3z2))))
144 formula(
145       forall([z1,x],
146              implies(R1(x,z1),
147                     forall([z2],
148                            implies(R3(z1,z2),
149                                   exists([z3],
150                                          and(R0(x,z3),R3(z3,z2))
```

```
                                    )
                                )
                            )
                        )
                ), I_B
).

% R0 is serial, transitive and euclidean (KD45: intention operator)
formula(forall([x], exists([y],R0(x,y))), I_serial).
formula(forall([x,y,z], or(not(R0(x,y)), not(R0(y,z)), R0(x,z))), I_transitive).
formula(forall([x,y,z], or(not(R0(x,y)), not(R0(x,z)), R0(y,z))), I_euclidean).

% R1 is serial, transitive and euclidean (KD45: belief operator)
formula(forall([x], exists([y],R1(x,y))), B_serial).
formula(forall([x,y,z], or(not(R1(x,y)), not(R1(y,z)), R1(x,z))), B_transitive).
formula(forall([x,y,z], or(not(R1(x,y)), not(R1(x,z)), R1(y,z))), B_euclidean).

% R2 is reflexive, symmetric and transitive (S5: historical necessity)
formula(forall([x], R2(x,x)), H_reflexive).
formula(forall([x,y], or(not(R2(x,y)), R2(y,x))), H_symmetric).
formula(forall([x,y,z], or(not(R2(x,y)), not(R2(y,z)), R2(x,z))), H_transitive).

end_of_list.

list_of_special_formulae(conjectures, EML).

% To proof:
% I[XSTIT]s /\ I[XSTIT](s->k) -> I[XSTIT]k
prop_formula(implies(
        and(box(I,box(XSTIT,s)),
        box(I,box(XSTIT,implies(s,k)))),
        box(I,box(XSTIT,k)))).

end_of_list.

list_of_settings(SPASS).
{*
  set_flag(DocProof,1).
*}
end_of_list.

end_problem.
```

B KI-frame

SPASS-file for standard KI-frame with to proof $I_a[a \text{ xstit}]k \rightarrow X(k)$

```
begin_problem(KI_frame).
list_of_descriptions.
  name({* KI_frame *}).
  author({* Nadine Hermans *}).
  status(unknown).
  description({*

        To proof: I[XSTIT]k -> Xk

  *}).
end_of_list.

list_of_symbols.
predicates[(k,0),
          (I,0), (K,0), (H,0), (XSTIT,0), (X,0),
          (R0,2), (R1,2), (R2,2), (R3,2), (R4,2)].
translpairs[(I,R0), (K,R1), (H,R2), (XSTIT,R3), (X,R4)].
end_of_list.

list_of_special_formulae(axioms, eml).

% Lin: -X-p <-> Xp
% <4>p <-> [4]p
% forall y1((-(xR4y1) \/ forall z1((xR4z1) -> (z1 = y1))) /\ exists z2(xR4z2))
formula(
       forall([y1,x],
              and(
                     or(not(R4(x,y1)),
                            forall([z1],
                                   implies(R4(x,z1),equal(z1,y1))
                            )
                     ),
                     exists([z2], R4(x,z2))
              )
       ), Lin
).

% Sett: []Xp ->[XSTIT]p
% [2][4]p -> [3]p
% forall z1((xR3z1) -> exists z2((xR2z2) /\ (z2R4z1)))
formula(
       forall([z1,x],
              implies(R3(x,z1),
                     exists([z2],
                            and(R2(x,z2),R4(z2,z1))
                     )
              )
       ), Sett
).
```

```
50
51  % XSett: [XSTIT]p -> X[]p
52  % [3]p -> [4][2]p
53  % forall z1((xR4z1) -> forall z2((z1R2z2) -> (xR3z2)))
54  formula(
55         forall([z1,x],
56                implies(R4(x,z1),
57                       forall([z2],
58                              implies(R2(z1,z2),R3(x,z2))
59                       )
60                )
61         ), XSett
62  ).
63
64  % KX: KXp -> K[XSTIT]p
65  % [1][4]p -> [1][3]p
66  % forall z1((xR1z1) -> forall z2((z1R3z2) -> exists z3((xR1z3) /\ (z3R4z2))))
67  formula(
68         forall([z1,x],
69                implies(R1(x,z1),
70                       forall([z2],
71                              implies(R3(z1,z2),
72                                     exists([z3],
73                                            and(R1(x,z3),R4(z3,z2))
74                                     )
75                              )
76                       )
77                )
78         ), KX
79  ).
80
81  % ER: K[XSTIT]p -> XKp
82  % [1][3]p -> [4][1]p
83  % forall z1((xR4z1) -> forall z2((z1R1z2) -> exists z3((xR1z3) /\ (z3R3z2))))
84  formula(
85         forall([z1,x],
86                implies(R4(x,z1),
87                       forall([z2],implies(R1(z1,z2),
88                              exists([z3],and(R1(x,z3),R3(z3,z2)))
89                              )
90                       )
91                )
92         ), ER
93  ).
94
95  % Unif-Str: <>K[XSTIT]p -> K<>[XSTIT]p
96  % <2>[1][3]p -> [1]<2>[3]p
97  % forall y1(-(xR2y1) \/ forall z1((xR1z1) -> exists z2((z1R2z2) /\ forall z3((z2R3z3)
        -> exists z4((y1R1z4) /\ (z4R3z3))))))
98  formula(
99         forall([y1,x],
100               or(not(R2(x,y1)),
101                      forall([z1],
```

```
                              implies(R1(x,z1),
                                     exists([z2],
                                            and(R2(z1,z2),
                                                   forall([z3],
                                                          implies(R3(z2,z3),
                                                                 exists([z4],
                                                                        and(R1(y1,z4),R3(z4,z3))
                                                                 )
                                                          )
                                                   )
                                            )
                                     )
                              )
                       )
                )
        ), Unif_Str
).

% K-S: []Kp <-> K[]p
% [2][1]p <-> [1][2]p
% (forall z1((xR1z1) -> forall z2((z1R2z2) -> exists z3((z3R1z2) /\ (xR2z3)))) /\
    forall z4((xR2z4) -> forall z5((z4R1z5) -> exists z6((xR1z6) /\ (z6R2z5)))))
formula(
        forall([z1,x],
                and(
                       implies(R1(x,z1),
                              forall([z2],
                                     implies(R2(z1,z2),
                                            exists([z3],
                                                   and(R1(z3,z2),R2(x,z3))
                                            )
                                     )
                              )
                ),
                forall([z4],
                       implies(R2(x,z4),
                              forall([z5],
                                     implies(R1(z4,z5),
                                            exists([z6],
                                                   and(R1(x,z6),R2(z6,z5))
                                            )
                                     )
                              )
                       )
                ))
        ), K_S
).


% X-Eff-I: []KXp -> I[XSTIT]p
% [2][1][4]p -> [0][3]p
% forall z1((xRz1) -> forall z2((z1R3z2) -> exists z3((z3R4z2) /\ exists z4((z4R1z3) /\
    (xR2z4)))))
```

```
formula(
       forall([z1,x],
              implies(R0(x,z1),
                     forall([z2],
                            implies(R3(z1,z2),
                                   exists([z3],
                                          and(R4(z3,z2),
                                                 exists([z4],
                                                        and(R1(z4,z3),R2(x,z4))
                                                 )
                                          )
                                   )
                            )
                     )
              )
       ), X_Eff_I
).

% (I => K): I[XSTIT]p -> K[XSTIT]p
% [0][3]p -> [1][3]p
% forall z1((xR1z1) -> forall z2((z1R3z2) -> exists z3((xRz3) /\ (z3R3z2))))
formula(
       forall([z1,x],
              implies(R1(x,z1),
                     forall([z2],
                            implies(R3(z1,z2),
                                   exists([z3],
                                          and(R0(x,z3),R3(z3,z2))
                                   )
                            )
                     )
              )
       ), I_K
).

% R0 is reflexive, symmetric and transitive (S5: intention operator)
formula(forall([x], R0(x,x)), I_reflexive).
formula(forall([x,y], or(not(R0(x,y)), R0(y,x))), I_symmetric).
formula(forall([x,y,z], or(not(R0(x,y)), not(R0(y,z)), R0(x,z))), I_transitive).

% R1 is reflexive, symmetric and transitive (S5: knowledge operator)
formula(forall([x], R1(x,x)), K_reflexive).
formula(forall([x,y], or(not(R1(x,y)), R1(y,x))), K_symmetric).
formula(forall([x,y,z], or(not(R1(x,y)), not(R1(y,z)), R1(x,z))), K_transitive).

% R2 is reflexive, symmetric and transitive (S5: historical necessity)
formula(forall([x], R2(x,x)), H_reflexive).
formula(forall([x,y], or(not(R2(x,y)), R2(y,x))), H_symmetric).
formula(forall([x,y,z], or(not(R2(x,y)), not(R2(y,z)), R2(x,z))), H_transitive).

end_of_list.

list_of_special_formulae(conjectures, EML).
```

```
206
207 % To proof:
208 % I[XSTIT]k -> Xk
209 prop_formula(implies(
210        box(I,box(XSTIT,k)),
211        box(X,k))).
212
213 end_of_list.
214
215 list_of_settings(SPASS).
216 {*
217   set_flag(DocProof,1).
218 *}
219 end_of_list.
220
221 end_problem.
```