社会结构中的形式脉络: 多主体交互的逻辑动态研究

申请清华大学-阿姆斯特丹大学联合授予
博士学位论文

李 大 柱

二〇二一年七月

Formal Threads in the Social Fabric: Studies in the Logical Dynamics of Multi-Agent Interaction

Dissertation Submitted to

Tsinghua University and University of Amsterdam

in partial fulfillment of the requirement

for a joint doctorate degree

by

Dazhu Li

July, 2021

Formal Threads in the Social Fabric Studies in the Logical Dynamics of Multi-Agent Interaction

ILLC Dissertation Series DS-2021-13

Institute for Logic, Language and Computation

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

We acknowledge the generous support of a 2-year Chinese Scholarship Council (CSC) scholarship.

Cover design by Dazhu Li.
Printed and bound by your printer.

Formal Threads in the Social Fabric

Studies in the Logical Dynamics of Multi-Agent Interaction

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in het auditorium van Tsinghua University

op donderdag 9 september 2021, te 08.30 uur CST

door Dazhu Li

geboren te Henan

Promotiecommissie

Promotores:	dr. A. Baltag	Universiteit van Amsterdam
	prof. dr. F. Liu	Tsinghua University
Copromotor:	prof. dr. J.F.A.K. van Benthem	Universiteit van Amsterdam
Overige leden:	prof. dr. F. Berto	Universiteit van Amsterdam
	prof. dr. Y. Venema	Universiteit van Amsterdam
	prof. dr. M.J.B. Stokhof	Universiteit van Amsterdam
	prof. dr. D. Zhu	Tsinghua University
	prof. dr. W. Wang	Tsinghua University
	prof. dr. H. Tang	Tsinghua University

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Dit proefschrift is tot stand gekomen binnen een samenwerkingsverband tussen de Universiteit van Amsterdam en Tsinghua University met als doel het behalen van een gezamenlijk doctoraat.
Het proefschrift is voorbereid in de Faculteit der Natuurwetenschappen, Wiskunde en Informatica van de Universiteit van Amsterdam en de afdeling Wijsbegeerte van Tsinghua University.

This thesis was prepared within the partnership between the University of Amsterdam and Tsinghua University with the purpose of obtaining a joint doctorate degree. The thesis was prepared in the Faculty of Science at the University of Amsterdam and in the Department of Philosophy at Tsinghua University.

学位论文指导小组、公开评阅人和答辩委员会名单

指导小组名单

刘奋荣	教授	清华大学
J.F.A.K. van Benthem	教授	阿姆斯特丹大学, 斯坦福大学, 清华大学
A. Baltag	副教授	阿姆斯特丹大学

公开评阅人名单

M.J.B. Stokhof	教授	阿姆斯特丹大学
Y. Venema	教授	阿姆斯特丹大学

答辩委员会名单

主席	M.J.B. Stokhof	教授	阿姆斯特丹大学
委员	F. Berto	教授	阿姆斯特丹大学
	Y. Venema	教授	阿姆斯特丹大学
	唐浩	教授	清华大学
	王巍	教授	清华大学
	朱东华	教授	清华大学

摘 要

社会生活的一个决定性特征是人与人之间的交互。我们的行为往往是对他人所做事情的反应，而别人再对我们的行为做出反应。这种无休止的纠缠在大量的场景中都有所体现，包括信息的交流、观点在社交网络中的传播、经济或学术活动中的竞争与合作，甚至社会关系自身都处在动态变化中。从社会学, 经济博弈论到社会认知论或行动哲学等很多学科都对这些现象进行了研究，本文从逻辑学的角度出发对其进行探索。社会互动是目前多主体系统逻辑中的一个核心主题，其牵扯哲学、计算机科学和人工智能等的交叉，由此产生的系统不仅进一步提高了我们对人类自身行为的理解，而且也被用来设计新的人类或人工主体的行为。本文是对多主体传统（特别是动态认知逻辑传统）的延续, 同时也探究了两个新的逻辑视角，它们凸显了社会交互中的两个进一步的特征。

第一个主题是处于不利条件下的多主体互动。其中，不同主体有着深层次的矛盾：他们尽力去改变交互发生的场景（通过物理的或其他的方式)。正如在一些场景中参与者发现自己受到敌对攻击时所发生的。为了对这些情形简洁清晰地建模，我们使用了特殊的“图博弈”，其中玩家可以在博弈过程中改变图形（即他们互动的场景)。在我们的核心博弈中，一个主体（“旅行者”）意图移动到目标区域, 然而另一个主体（“破坏者”）尽可能地去删除图中的链接以阻止旅行者。这些图博弈特别适合于从逻辑的角度进行分析。通过对现有文献进行扩展，我们对一类图博弈提供了一个完整的逻辑分析：其中，旅行者当前位置的链接按照某种可定义的方式被删除。这种“基于一个描述的局部破坏”涵盖了许许多多的场景，并支持一个丰富的逻辑理论。

虽然上述场景看起来可能有点“负面”，但是链接删除作为一个抽象的技术也可能是有益的：我们通过下一步关于教、学中主体间的互动来表明这一点。为此, 我们考虑一个更真实的具体场景，并设计了一个更丰富的图博弈，其中，老师对链接的删除可以表示两种指正：指出学习者已经犯下的错误，或引导学习者远离未来可能出现的错误。同样的，我们提供了一个逻辑语言去分析这些场景，并展示这如何为分析学习中的动态提供一个丰富的框架，其比形式化学习理论中的标准场景更深入地分析了过程上的细节。

如果我们将逻辑方法在以上研究中的作用看作是对社会场景提供更为精确和详细的分析，那么论文的第二部分对逻辑的使用多多少少是朝向另一方向的，即去寻找不同场景中抽象的一般结构。在这里我们特别感兴趣的是社会活动中主体

行为上的依赖性。首先，我们从抽象角度探讨了多主体系统中时间上的动态依赖概念。为了刻画社会交互的时间维度，我们研究了一个能够体现动态系统中函数依赖性的核心推理的逻辑系统。这需要我们用时态逻辑中的工具去扩展现有的关于依赖性的模态逻辑，由此我们提出了一个新的关于动态系统中的行为和依赖性的逻辑。针对这一逻辑体系，我们证明了它的完全性结果和一些其他的性质。此外，由于文献中关于动态系统的使用大都涉及到状态空间上的一个拓扑，我们也提供了这样一个扩展，其可以描述这样一种社会交互情形的特征：其中，我们只有不精确的（即使是可改善的）方式去测量相关的变量。这带来的是一个更为丰富的逻辑，我们称之为动态连续性依赖。一个看待这些系统的方式是将其作为对当前进化博弈论中关于社会行为分析的一个概况抽象。

接下来，正如我们的第一部分那样，在研究了抽象的基本理论之后，我们进而还考虑了更现实的社会场景。论文研究的具体案例是群体中观点或行为的传播。其中，主体基于周围其他主体的行为来按照自己的阈值更新自身的行为。在这项工作中，我们强调了信息的关键作用，并提出了一个逻辑来体现如何对我们一贯以来的分析添加一个认知维度。

最后，我们对全文进行总结，并提出一些我们的分析所带来的新的问题：这不仅包括多主体系统逻辑中的技术方面的开问题，也包括我们应该如何把社会实体放在起始位置的概念方面的反思。

关键词：社会交互；图博弈；动态多主体逻辑；依赖性；动态系统；行为传播

Samenvatting

Interactie is een centraal aspect van het sociale leven. Onze handelingen reageren doorgaans op wat naderen hebben gedaan, en anderen reageren weer op ons. Dee voortdurende vervlechting komt in veel sociale situaties voor, van uitwisselen van informatie tot het ontwikkelen van publieke opinie, samenwerking en tegenwerking in economische of academische activiteiten, en zelfs sociale relaties zelf zijn voortdurend onderhevig aan verandering. Hoewel deze verschijnselen al uitvoerig zijn bestudeerd in disciplines zoals sociologie, economische speltheorie, of sociale epistemologie en filosofie van het handelen, concentreert dit proefschrift zich op een logisch perspectief. Sociale interactie is een belangrijk onderwerp in moderne logische analyses van meer-actor systemen, en de inzichten die op deze manier zijn verworven worden toegepast in het begrijpen van bestaand sociaal gedrag, maar ook het ontwerpen van nieuwe vormen van gedrag van mensen en machines. Voortgaand op deze traditie, vooral in zijn dynamisch-epistemische varianten, onderzoeken wij twee verdere belangrijke aspecten van sociale interactie.

Ons eerste onderwerp is sociale interactie onder ongunstige omstandigheden. Dit komt voor wanneer actoren diametraal tegenover elkaar staan, en zelfs de (fysieke) omgeving trachten te veranderen waarin hun interactie zich afspeelt, bijvoorbeeld in scenario's met een vijandelijke aanval. Om zulke scenario's helder te modelleren gebruiken we een speciaal soort 'graafspelen' waar spelers de graafstructuur van punten en verbindingen, die dient als hun speelveld, veranderen tijdens het spel. In ons centrale genre graafspelen wil een Reiziger een bepaald doelgebied bereiken, terwijl een demon dit zoveel mogelijk tegenwerkt door verbindingen in de graaf te verwijderen. Zulke spelen lenen zich heel goed voor logische analyse, en voortgaand op eerdere literatuur, geven we een volledige analyse van geldig redeneren over graafspelen waar de Demon verbindingen kan weghalen bij de huidige positie van de Reiziger, volgens een voorhanden zijnde beschrijving. Dit beeld van 'lokale sabotage onder een beschrijving' past op vele sociale scenario's, en leidt to een rijke logische theorie van redeneren.

Hoewel deze formulering 'negatief' kan klinken, kan weghalen van verbindingen evengoed gunstig zijn: we bestuderen ook scenario's van leren en onderwijzen, waar dit het geval is. In een meer concreet realistisch scenario haalt de Leraar verbindingen weg die corresponderen met foute redeneerstappen die de Leerling heeft gemaakt, of verbindingen

die later tot dwaalwegen en fouten zouden leiden. Weer geven we een precieze logische taal voor zulke scenario's, en we laten zien hoe het resultaat een rijk model oplevert voor de procedurele dynamiek van leren en onderwijzen dat een natuurlijke aanvulling vormt op de bestaande formele leertheorie.

In de onderwerpen tot nut toe verschaffen logische methoden meet detail en precisie in de analyse van sociale scenario's. In het volgende deel van dit proefschrift keren we de richting om: logische methoden helpen nu om algemene abstracte structuren te vinden die door veel concrete scenario's heen spelen. Onze speciale interesse is de notie van afhankelijkheid van gedrag voor actoren in sociale activiteiten.

Om te beginnen onderzoeken we de abstracte notie van afhankelijkheid door de tijd heen in meer-actor systemen. Om zicht te krijgen op de temporele dimensie van afhankelijk sociaal gedrag, ontwikkelen we een logisch systeem dat fundamenteel redeneren over functionele afhankelijkheid beschrijft die zich manifesteert in het verloop van tijd. Hiervoor gebruiken we een combinatie van bestaande modale logica's van afhankelijkheid met begrippen uit de temporele logica, en het resultaat is een logica van handelen en afhankelijkheid in dynamische systemen waarvoor we volledigheid en andere eigenschappen bewijzen. Daarenboven, omdat de meeste dynamische systemen in de literatuur een topologie hebben op hun toestandsruimte, geven we ook een topologische versie van ons systeem die informatie kan beschrijven over sociale interactie wanneer we alleen niet-exacte (hoewel verfijnbare) manieren hebben om de relevante variabelen te meten.Het resultaat is een volledige logica van wat we dynamische continue afhankelijkheid noemen. Dit systeem kan onder meer worden beschouwd als een veralgemening van huidige analyses van sociaal gedrag in de evolutionaire speltheorie.

Vervolgens, net als in ons eerste deel, testen we de ontwikkelde abstracte basistheorie in een meer realistisch sociaal scenario. We beschouwen in het bijzonder de ontwikkeling en verspreiding van meningen of gedrag in sociale gemeenschappen, waar actoren hun gedrag bijstellen aan de hand van gedrag dat zij hebben geobserveerd by hun buren in het sociale netwerk, volgens een regel gebaseerd op een drempel van activatie. We benadrukken de essentiële rol van informatie in dit proces, en geven een logische analyse van de wat noddig is om een epistemische dimensie toe te voegen aan de stijl van analyse in dit proefschrift.

In een concluderend hoofdstuk maken we de balans op van wat er is bereikt, en wat zich voor nieuwe taken voordoen. We identificeren vele technische open problemen in de

logische studie van meer-actoren systemen in onze stijl, maar ook mogelijke conceptuele heranalyse van hoe men sociale entiteiten het best kan modelleren.

Trefwoorden: sociale interactie; graafspelen; dynamische meer-actor logica's; afhankelijkheid; dynamische systemen; evolutie van gedrag

Abstract

Interactions between people are a defining feature of social life. Our actions tend to be reactions to what others have done, while others again respond to our behavior. This never-ending entanglement can be observed across a wide range of settings including exchange of information, spread of opinions in social networks, cooperation and competition in economic or academic activities, and even social relationships themselves are in dynamic flux. While these phenomena have been studied in many disciplines, from sociology and economic game theory to social epistemology or philosophy of action, this dissertation pursues a logical perspective. Social interaction is a core topic in current logics of multi-agent systems at the interface of philosophy, computer science and AI, and the resulting systems have been applied to better understand human behavior, but also to design new forms of behavior by both human and artificial agents. This dissertation continues within this multi-agency tradition, especially, that of dynamic-epistemic logics, and explores two new logical perspectives that highlight two further basic properties of social interaction.

The first topic is multi-agent interaction under adverse circumstances. This arises when agents are deeply at odds, to the extent that they try to change the very setting (physical or otherwise) where their interactions take place — as happens, for instance, when actors in some standard scenario find themselves under hostile attack. For a crisp modeling of such scenarios, we use special 'graph games' where players can change the graph, i.e., their playground, during play. In our central game, one agent (the Traveler) wants to reach some region representing a goal, while the other player (the Demon) obstructs the Traveler as much as possible by removing edges from the graph. These graph games turn out to be highly amenable to logical analysis, and extending existing literature on these scenarios, we provide a complete logical analysis of graph games where obstruction consists in removing edges at the current position of the Traveler in some definable manner. This 'local sabotage under a description' covers many scenarios and supports a rich logical theory of valid reasoning.

Although the above scenario may look 'negative', edge removal as an abstract technique can also be beneficial: we demonstrate this by next studying the interactions of agents engaged in learning and teaching. For this purpose, we consider a more realistic concrete scenario, and design richer graph games where edge removals by a Teacher rep-

resent corrections of two kinds: pointing out errors already made by a Learner, or steering the Learner away from potential future mistakes. Again we provide a logical language for analyzing these scenarios, and we show how this provides a rich framework for analyzing the dynamics of learning that goes into more procedural details than standard scenarios in formal learning theory.

One can view the role of logical methods in the preceding cases as providing more precision and detail in the analysis of social scenarios. The second part of the thesis uses logic in more or less the opposite direction: finding abstract general structures that play across many scenarios at the same time. Our particular interest here is the notion of dependence of behavior for agents engaged in social activities.

First, we explore the abstract notion of dynamic dependence over time in multi-agent systems. To capture the temporal dimension of social interaction, we develop a logical system embodying the core reasoning about functional dependence in dynamical systems. This requires extending existing modal logics of dependence with devices from temporal logic, and the result is a logic of action and dependence in dynamical systems, for which we show completeness and other properties. Moreover, since most uses of dynamical systems in the literature involve a topology on the state space, we also offer an enrichment. We introduce a topological version of the system that can describe information about social interaction when we have only imprecise (though refinable) ways of measuring the relevant variables. The result is a richer logic of what we call dynamic continuous dependence. One way of viewing these systems is as a generalization of current analyses of social behavior in evolutionary game theory.

Next, as in our first part, having developed the abstract base theory, we consider what else needs to come in to deal with more realistic social scenarios. Our case study is that of diffusion of opinions or behaviors in communities, where agents update their behavior based on what their neighbors in the social network do, according to some threshold rule. We highlight the crucial role of information in making this work, and present a logical case study of what it takes to add an epistemic dimension to our style of analysis so far.

We conclude by taking stock, and pointing at the many new issues raised by our analysis. These include many technical open problems in the logic of multi-agent systems, but also conceptual rethinking of how one should represent social entities in the first place.

Keywords: social interaction; graph games; dynamic multi-agent logics; dependence; dynamical systems; evolution of behaviors

Contents

Chapter 1 Introduction

1.1 Motivations

Interactions between people are a defining feature of social life. From the start of our lives, our actions tend to be reactions to what others have done, while others again respond to our actions, in an endless cycle. This entanglement of behavior can be observed across a wide range of settings people find themselves in, such as exchange of information, spread of opinions in social networks by assent or dissent, cooperation and competition in economic (or academic) activities, and even social relationships themselves are in dynamic flux. While these phenomena have been studied in a wide range of disciplines, from sociology and economic game theory to social epistemology or philosophy of action, there is also room for a logical perspective. Social interaction is a core topic in current logics of multi-agent systems at the interface of philosophy, computer science and AI, and the resulting systems have been applied to better understand human behavior, but also to design new forms of behavior by both human and artificial agents (Shoham and Leyton-Brown, 2008; Wooldridge, 2002).

Dynamic logics of knowledge, belief, and social structures. One important line in this field which is congenial to this dissertation is logical studies of dynamic phenomena in social interaction. There are dynamic-epistemic logics for studying exchange of information among different agents, including public announcement (Plaza, 1989), private or semi-private communication (e.g., Baltag et al., 1998; van Benthem et al., 2006; Wang et al., 2010), and various forms of group knowledge and belief (e.g., Baltag et al., 2018), as well as epistemic-temporal logics of message passing: Fagin et al. (1995); Parikh and Ramanujam (2003). While these logics tend to focus on informative single agent interactions, another line of inquiry has focused on group phenomena over time, such as influence on individual beliefs through social relationships (Liu et al., 2014; Seligman et al., 2011), peer pressure (Liang and Seligman, 2011), or diffusion of opinions in communities (Baltag et al., 2019b; Christoff and Hansen, 2015; Shi, 2021). Finally, there are also dynamic logics for analyzing the evolution of social structures among agents, such as creation of communities (Smets and Velázquez-Quesada, 2020), or structural changes in the presence

of social conflicts (Pedersen and Slavkovik, 2017).

This tradition of dynamic logics forms the backdrop to the topics in this thesis.[1] Continuing in its spirit, we will explore two further basic properties of social life that have received less attention so far in the logical literature. The first is multi-agent interactions under adverse informational, and even physical, circumstances that are common in real life. Our second topic are the bonds created by social interactions: dynamic dependencies unfolding over time between social actors. We now consider each of these in more detail.

Graph games for changing social and physical environments. Social interactions usually take place in a physical environment, and players' goals may be so antagonistic that they engage in various forms of obstruction and sabotage changing the environment. This happens, for instance, when actors in some standard scenario find themselves under hostile attack (a very common phenomenon in ancient societies, or in modern internet systems). And even without deliberate destructive actions by all or some players, such drastic changes may occur when a social system has to function in a hostile or malfunctioning environment. We hasten to add that there are also benign examples of all this, such as teachers guiding their students toward desirable learning goals by removing distractions and temptations.

To understand these realistic scenarios, we analyze what happens when agents are deeply at odds, to the extent that they try to change the very setting (physical or otherwise) where their interactions take place. For a crisp modeling of the abstract essence of such scenarios, we will use a special sort of *graph games*, a widely used technique in computational logic and graph theory (van Benthem and Liu, 2020). In the games to be introduced below, players can change the graph, i.e., their playground, during play. In our central game, one agent (the Traveler) wants to reach some region representing a goal, while the other player (the Demon) obstructs the Traveler as much as possible. These graph games are known to be highly amenable to logical analysis, and extending existing literature on these scenarios, we provide a complete logical analysis of graph games where obstruction consists in removing edges at the current position of the Traveler in some definable manner. This 'local sabotage under a description' covers many social scenarios and supports a rich logical theory of valid reasoning.

1 The general logical literature on multi-agent systems is much wider than this, and indeed too wide to survey here, The reader can form an impression of its range by looking at books like Hendricks and Hansen (2016); Shoham and Leyton-Brown (2008); van Benthem (2014); Wooldridge (2002).

More concretely, we start from the paradigm of '*sabotage games*' (van Benthem, 2014), which highlight essential features of competition with potentially disruptive opponents, while still maintaining tight connections with logic. Here is the classical version:

Sabotage game. A sabotage game is played on a graph, representing the environment, with a starting-node and a goal-node, or more generally, a goal-region: in each round, a player Demon first cuts a link anywhere in the graph [so, Demon acts globally], and then the other player Traveler moves along an edge that is still available where she stands [Traveler acts locally]. Traveler wins if she arrives at a node in the goal region: if this does not happen, and no more moves are possible, Demon wins.[1]

As an important property of these scenarios, it is easy to see that sabotage games are *determined* in the sense of game theory, at each position, one of the two players, Traveler or Demon, has a winning strategy. The reason is that Zermelo's Theorem applies (Osborne and Rubinstein, 1994), the games are two-player zero-sum with perfect information, and also, there is a fixed finite horizon to when the game is over, since Demon has only finitely many links to cut from.

Although sabotage games may look very simple, they apply in principle to a wide range of interactive scenarios with adverse circumstances. They also fit well with many actual parlor games where blocking can be done by devices such as putting pawns on certain positions to make them inaccessible, and so on (van Benthem and Liu, 2020). This is no coincidence, the analogy between games people play and the game structure in serious social activities has often been noticed (e.g., Franklin, 1786; Huizinga, 1949).

However, when we take sabotage to concrete social settings, we often find that more structure needs to be accounted for. Here is an illustration that may speak best to readers familiar with crime series (see Figure 1.1):

Example 1.1: Policewoman Alice is on a mission with her colleague Bob, driving from point i to regions t and g which are the two locations of a criminal gang. However, Alice does not know that Bob is in fact a corrupt cop who works for that gang. To hinder Alice, Bob secretly keeps sending their positions to the criminals, while Alice is focused on driving and does not realize what is happening. As a response to Bob's messages, the group takes action: with the information on Alice's current position in hand, they block one or more roads (represented by links in the graph) that she may take soon, to guide

1 See Section 1.4.1 for more on sabotage games and its matching system of *sabotage modal logic*.

her away from the target locations, or even stop her as early as possible. In this setting, to reach her destination, what Alice needs to do is to handle traffic conditions that keep changing.

A road map of a highly simplified situation of this sort might look as follows.

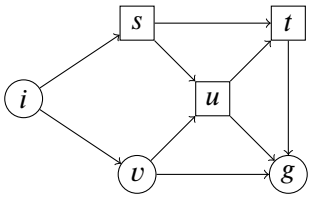


Figure 1.1 Alice's driving map. Nodes stand for the main locations that she may go through, directed arrows represent roads as well as her driving direction, and the two kinds of shape, square and circle, denote different relevant properties of locations.

For simplicity, this scenario can be taken as a game played by two players: Alice and Bob, where we see the gang just as a tool for Bob to use.[1] As in the original sabotage game, Alice aims to reach her goal region, while Bob tries to stop Alice by blocking roads (or more technically, removing links). However, compared with the Demon in sabotage games, Bob's behavior is now much more target-directed: to stop Alice as early as possible, he deletes links connected to the current position of Alice. This is a very common situation, both social actors are localized now, and indeed, they are at the same location. This removal of the asymmetry of global Demon and local Traveler in the original sabotage games has been studied by several authors (e.g., Aucher et al., 2018; Löding and Rohde, 2003b), and it simplifies the game analysis. In particular, Zhang (2020) has shown that while the solution complexity of sabotage games with a global Demon is PSPACE-complete, solving local sabotage games for who has the wining strategy at any given point only takes polynomial time, a significant jump downward in computational complexity and intuitive difficulty. Even so, there is an issue of whether reasoning about local sabotage games is easier than that for the original scenarios, and this will in fact be one of our guiding questions below.

However, the Alice-Bob scenario is still underspecified. Bob's situation may actually both better and worse than the above description may have suggested. On the one hand, the gang may be able to block *more than one* road at the same time, instead of the single roads allowed by the original game.[2] On the other hand, choices may not be arbitrary,

1 Of course, the gang might have goals that are slightly different from Bob's. More-agent sabotage games are an interesting extension of the sabotage set-up that has not been studied systematically so far.

2 The effect of progressively allowing Demon more removals in one round has not been studied systematically, but

but may have to be guided by a *description*: the gang may only be able to block all roads satisfying a certain description. For instance, in blocking your access to websites in a network, I may have to do different blocking actions depending on what sort of website I want to block.

Concretely, in the above example, the two properties 'diamond' and 'circle' represent two definable choices of sabotage actions. If we assume that Bob has to delete all nodes satisfying one of these properties, the strategic situation is as follows. If Bob chooses node s, Alice will move to v, and Bob must block the link to g, after which Alice moves to u. Here Bob can only block one link to a goal point, and Alice can get to the goal region via the remaining link. However, if Bob starts by blocking v, Alice will move to s, and here, Bob can delete all square successor nodes, trapping Alice before she ever reaches the goal. Thus, Bob has a winning strategy with Alice at the initial point. In fact, this analysis tells us who has the winning strategy at each location for Alice in the original graph.[1]

Questions for logical analysis. Looking at sabotage-style graph games from the perspective of logic, our primary concern is the structure and complexity of reasoning about them. Players have to reason inside the game about effects of their actions, but as external observers, we also want to reason about properties of one game, or classes of graph games. For this, we need to introduce logical syntax, and this is what we shall do in this thesis.

It was already observed in the original paper (van Benthem, 2005) that sabotage games suggest an unusual kind of modal logic, where ordinary existential modalities $\Diamond\varphi$ represent available steps by Traveler to accessible points satisfying φ, whereas a new sort of existential 'deletion modality' $\blacklozenge\varphi$ states that some link can be cut from the graph so as to make φ true at the current point. This modal logic was studied in (Löding and Rohde, 2003a) who showed that, despite the apparent simplicity of this modal logic, it is undecidable. This is surprising since modal logics of ordinary graph games tend to be decidable (van Benthem and Liu, 2020): reasoning correctly about social scenarios with environmental change comes at a price.[2] Also, while sabotage modal logic is axiomatizable in principle (this follows from the first-order translation in van Benthem (2005)), a perspicuous Hilbert-style axiomatization has long been an open problem, and the best available

we will leave this technical issue aside in what follows.

1 Again, one should appreciate the generality of the scenarios sketched here. One might also implement all of the above as a real or parlor game where players can manipulate the color of traffic lights, and so on.

2 The complexity may have to do with the fact that 'modal sabotage logic' can express much more than just basic game properties (for more on this, see Aucher et al., 2018).

result so far seems the recent axiomatization using some small expressive extensions to the basic language in (van Benthem et al., 2021a).

So, here is our first question: *What is the complexity of reasoning about local sabotage, and how can we axiomatize it?* This question immediately generalizes to local sabotage with definable link deletion, a topic that had not been studied in the earlier literature. We will find answers to these questions in Chapter 2 of this dissertation, plus a lot more information of a theoretical nature that comes with the study of logical languages and logical systems. The interest of a logical system is not just whether it answers some initial question: once in existence, it can also generate further motivations by its very nature.

A positive scenario: modeling learning through graph change. Although the sabotage view of social interactions may look a bit 'negative', abstract link modifications do not have any negative nature at all. They could just as well represent the action of a gentle guiding hand of a parent or teacher removing false paths from the environment of a child.

In particular, the original sabotage game has been interpreted as a scenario of *learning and teaching* in (Gierasimczuk et al., 2009). In this reading, points and edges in graphs stand for different hypotheses and possible inferences conjectured by Learner. A transition from point a to another b represents Learner's inferring b from a, while removing links is regarded as Teacher's feedback: helping Learner to eliminate incorrect inferences. While this is appealing to some extent, and the formal results obtained were suggestive, this interpretation will only become more plausible as a contribution to formal learning theory if we confront it with more detailed scenarios.

Here is an example that readers of this dissertation may recognize, having served in one or more of the roles involved:

Example 1.2: After checking a proof written by Learner (L), Teacher (T) begins to talk:

T: "You did not prove the theorem yet."

L: "Why? I started with the axioms, showed intermediate lemmas step by step, and finally reached the statement of the theorem."

T: "Your final step to show the theorem that is the goal is correct, but you in fact arrived there by accident, as the inference from lemma α to lemma β in your proof is wrong."

L: "Oops! I see. Then, my steps after β do not make sense. But, how about a new

lemma proving γ from α? Now I think I can get to the theorem."

T: "Alas, γ cannot be inferred from α either."

L: "Sorry."

T: "No problem: it is just a potential mistake. But actually, you miss another lemma δ that can be derived from α. I believe you might be able to show the theorem with it."

L: "Thanks! You are right! Now I am going to search for a correct proof with δ."

This short episode already goes beyond the learning/teaching scenario in (Gierasimczuk et al., 2009), and it raises several interesting issues.

One is that there may be *several kinds of mistakes* that need to be addressed: actual mistakes made, and potential mistakes to be avoided. And this again calls for a rethinking of sabotage-style scenarios. The latter were 'history-free' in that Demon acts only on the current location of Traveler: how Traveler arrived there is not relevant. But in the present scenario, *the history matters*. Teacher's removing mistakes that were actually made acts on the history so far (and makes all further moves on that history suspect), while eliminating potential mistakes affects the future from the current point.[1]

Making this precise again leads to many further issues. We might stipulate that Teacher's pointing out an incorrect step removes the whole actual history after that step, resetting Learner to the last point before the mistake. Also, the Teacher may point a Learner to facts that were ignored, and also, Teacher may point at correct inferences, i.e., links that should not be cut. In terms of game design, this calls for a more powerful Teacher: in addition to removing links for wrong transitions, Teacher is also capable of *adding links* to graphs. Moreover, the winning conditions may be more complex than in the original sabotage game. Learner need not win when the goal region is reached (a history-free condition), but only when that goal region has been reached in the right way.[2]

Thus we have a new set of questions. *How can we model realistic Learner/Teacher scenarios of the above kind as graph games?* And once these are in place, *What is a correct modified logic of graph games for modeling such learning/teaching scenarios?* Answers to these questions will be found in Chapter 3 of this dissertation. It presents a logical

1 Given the abstract nature of graph games, it is technically possible to absorb histories into point of a *new graph* which then looks history-free, but this would lose us the intuitive phenomena we are after here, so we will not explore this style of remodeling.

2 In some original interpretations, the Learner was supposed to maximally seeking problematic escapes into ignorance and empty pleasures, while the Teacher sought to guide the student to some educationally desirable region. But in our new version, the goals of Learner and Teacher may well be aligned.

analysis of teaching/learning scenarios based on suitably modified and enriched graph games. This chapter may be seen as an more concrete application-inspired counterpart to the more theoretical investigation in Chapter 2.

Zooming out to more global logical structures in interaction. Next, having shown the power of graph games for changing social environments, and investigated their properties in matching logical systems, we also encounter a potential problem. Generally speaking, as we shall see, the logics that we found in the first two chapters, though expressive and capable of formulating many kinds of social scenarios, are complex, indeed *undecidable* systems of reasoning. Now we may mitigate this by perhaps restricting attention to just those fragments consisting of the precise formulas in our modal logics of graph change that express properties of graph games in a narrower sense. A better approach might be to identify the general sources of the undecidability, where an interesting perspective is the amount of *memory use* identified as a costly computational device in (Areces et al., 2011). However, the first approach is somewhat ad-hoc, and the second too computationally oriented for our purposes.

Instead, we propose to move away from specifics of computation in graph games by 'zooming out' to more global structures found in social interaction. While our graph logics went into quite some detail of what happens during a game, 'zooming in' on the game board and the game dynamics, we can also use logical tools to find reasoning patterns for more global notions driving social interaction.[1]

Dependent behavior in social settings and finding its minimal logic. Perhaps the most basic notion in extensive games over time that creates a social 'bond' is that of *dependence* of actions. Once I commit to a strategy, my actions will come to depend on yours, since my strategy prescribes a response. Thus, games create dependence patterns among players that may be considered an essential aspect of social life.

Not surprisingly, logicians have long recognized the importance of dependence, not just in games, but widely across the sciences and daily life. Pioneering contributions were made by Hintikka (1973), whose 'game-theoretic semantics' stresses the dependence patterns created by logical $\forall\exists$ quantifier combinations. For more recent versions of logics of dependence, and independence, cf. Hintikka and Sandu (1997); Väänänen (2007). However, as they stand, these logics are not suitable for our purposes, since they tend to

1 For a similar approach extracting general social postulates from the specifics of game theory, see (Johansen, 1982).

have high (second-order) complexity.[1] But the reason for our departure from the preceding logics was precisely that we want to lower complexity, by dropping specific features of earlier mathematical models, in the 'content vs. wrappings' spirit of van Benthem (1996).

Accordingly, we make our tools much weaker, taking our point of departure in a recent low-complexity modal approach to reasoning about dependence, put forward in (Baltag and van Benthem, 2021b). Here functional dependence between variables (that can typically stand for social actors) is defined in a minimal semantic manner, leading to a decidable core logic of dependence, on top of which the proof-theoretic surplus of specific application areas where dependence plays a role can be determined from case to case.

Yet, while this is a good start, strategic social interaction has one feature that is not covered by these systems: namely, dependence *over time*. We illustrate this with one more simple graph game.

Example 1.3: On the graph depicted in Figure 1.2, two players, Xanto and Ora, are playing a game of occupying territories. In each round, Xanto first labels a node with a letter '**X**' and then Ora labels a node with '**O**', thus 'occupying' these nodes. Here, a node can be occupied by a player only if it has not been occupied by anyone yet, and it is not adjacent to a node occupied by the other player in the previous round.[2] A player wins as soon as the other player cannot legally find a node to occupy.

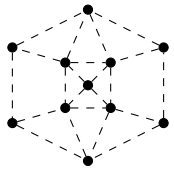


Figure 1.2 Occupying territories.

In this concrete situation, it is easy to see that Xanto has a winning strategy. More generally, on all finite centrosymmetric graphs with a centre, the starting player has a winning strategy: "first occupy the centre, then in subsequent rounds, just occupy the node centrosymmetric to that chosen by the other player in the previous step".[3] Of course, as with the earlier graph games, this abstract geometrical game can be given many concrete interpretations, from warfare to peacefully laying claims to resources.

1 These two approaches also emphasize non-classical features of reasoning about dependence which we think are orthogonal to the phenomenon of strategic dependence per se.

2 The adjacent relation is represented by the dashed links.

3 For instance, in the above example, after Xanto occupies the centre, if Ora chooses to occupy the node on the top left, then Xanto can occupy the node on the bottom right.

What we see in the occupying scenario is *dynamic dependence* with a time delay: the winning strategy involves Xanto's move dependent on that by Ora in the previous round.[1] This is again characteristic for social interactions: stable dependencies unfold over time, with Tit-for-Tat as a prime example: I copy what you did in the previous round, a strategy that has been claimed to play a crucial role in the very emergence of social cooperation (Axelrod, 1984).

Even so, all existing dependence logics that we know of ignore this temporal structure, and hence we arrive at a next set of questions. *How can we model dynamic dependence over time?* And as an associated issue of reasoning: *Is the logic of dynamic dependence still of the same simplicity as that for static dependencies?* We will answer these questions in Chapter 4, providing a precise model, and determining its complete logic.[2] As for the complexity of this basic style of reasoning, we suspect that it is decidable, but we only offer a decidability proof for a (considerable) fragment of the full language.

One general abstract way of thinking about our analysis of strategic dependence over time in extensive games is in terms of *dynamical systems* where a number of agents change state simultaneously in each step, according to some transition function representing a joint strategy.[3] Dependencies then show as correlations between actions of agents across transition steps. Dynamical systems have a wide range of application in the physical sciences, but they also underlie information-driven social processes (Klein and Rendsvig, 2017; Shi, 2021).[4]

But dynamical systems also suggest a desirable extension of our analysis so far. They typically come with an idea of 'closeness': is the transition function intuitively continuous, in the sense that small changes lead to small effects — or can there be discontinuous jumps in social behavior? Mathematically, dynamical systems often come with a *topology* on their state space, and this topological perspective fits well with recent trends in mathematical epistemology where topologically open sets represent results of possible measurements that we can take on the state of the system. This considerably extends the scope of our analysis to include empirical variables that cannot be measured precisely, but

1 Dynamic dependence could also have been illustrated with strategic behavior in our earlier sabotage game, but we added the occupation scenario to show the variety of simple, yet non-trivial graph games.

2 It should be noted that our analysis is based on strategies in extensive games. The basic modal dependence logic of Baltag and van Benthem (2021b) has also been applied to games in strategic form as a way of analyzing collective agency in (Shi and Wang, 2021).

3 For the preceding sequential games, we can let the non-active players perform an identity action.

4 Dynamical systems also underlie computational paradigms such as cellular automata, see Berto and Tagliabue (2021).

only approximated by successive measurements. (This principled approximate measurability may hold for physical properties, but it also applies to mental states such as anger or approval in other agents.) All this leads to further issues: *Is there a topological extension of our dynamic dependence logic, and is it still of the low complexity that we found for the non-topological case?* The first questions will be answered in the affirmative in Chapter 4 below. However, w have not yet managed to settle the second issue of decidability, i.e., the precise complexity of this style of reasoning.

Dynamical systems for opinion formation: a case study in adding knowledge. With the use of dynamical systems, we also approach a significant divide in current logical studies of agency (e.g., Skyrms, 1990). Traditional agent systems, and the games we have presented so far, typically assume 'high rationality': agents are aware of their situation and actively seek best responses, based on information and reasoning. But there is also 'low rationality: the behavior of preprogrammed agents who just follow some hard-wired update rule, for instance, adopting the opinions of the majority of one's neighbors in a graph modeling a social structure with relations such as physical proximity, communicative connectedness, or plain friendship. In this interpretation, agents no longer play on a given graph: they are themselves nodes in the graph, but these nodes can change their properties according to some update rule. The contrast, or the interface, between high and low rationality is of broad significance to understanding human social behavior, or interactions between human and artificial agents (van Benthem et al., 2021b).

Our final topic in this thesis lies in this area, though it only addresses one crucial feature of rational agency that manifests itself in both high- and low-rationality settings, though in different guises. First, as for low rationality, we already mentioned a growing logical literature on opinion formation over time by communities of agents following fixed rules of a dynamical system, i.e, hard-wired behavioral strategies, and the long-term patterns that may come out of such systems. In particular, there is a body of work using the same techniques as in this dissertation to study so-called 'threshold models' from the social sciences, where behavior gets adopted once the number of neighbors adopting it passes some threshold (Easley and Kleinberg, 2010). Containing the dynamic logic-style analysis in Baltag et al. (2019b); Christoff (2016); Liang and Seligman (2011); Liu et al. (2014), our final offering is an extension of such models to cover more realistic aspects of opinion formation, in the same spirit of moving toward greater realism as in our earlier topics.

Instead of giving a concrete scenario, we merely describe the crucial dimension that we have in mind. In all our examples so far, we made a tacit assumption that the agents have complete *knowledge* of their total environment and the actions of other players. In the sabotage game, Demon can see where Traveler is located, while Traveler can see the links cut by Demon. In the teaching game, Teacher and Learner are aware of the relevant network of claims and implications. But in our discussion of dependence, knowledge did start entering. First, the variables in a dependence modeling support a possible interpretation as agents, and in that case, dependencies between variables represent informativeness relations between agents (Baltag and van Benthem, 2021b). And also, our topological extension provided a setting where knowledge about exact values of variables can be improved by means of measurements. This suggests adding an explicit epistemic dimension to all of our previous topics, in terms of what (highly-)rational agents *know*, and can (deliberately) *observe* about their environment. While this particular direction would go far beyond the scope of this dissertation (a few thoughts can be found in the Conclusion chapter, in terms of dynamic-epistemic logics and extensive games with imperfect information), the dynamical systems setting presented here does support a simpler exploration in the low-rationality realm.

Low-rationality agents will not necessarily (perhaps even: necessarily not) reflect on what they do, but still, their actions can depend crucially, not on mere physical facts but on *information* about these facts. Instead of offering further concrete scenarios, we will just outline the reasons for going this way. Even a robot following a deterministic rule responding to its environment (say, "pick up a rock from the Mars surface if there is one in reach of my grabbing arm") must have sensors giving it information about the presence of that rock: it acts on the information, not the mere physical fact. But the same happens in the social scenarios that we are interested in. Consider a social group with friendship relations where agents will follow the opinion of the majority of their friends. Suppose also that friendship implies the possibility of communication, and hence of coming to know the opinions of one's friends. We might say then that a threshold rule like 'adopt the view of the majority of one's friends' is hardly usable if one does not know those opinions. What is called for is rather *epistemized update rules* such as 'adopt the majority view among one's friends if one *knows* what that view is', or, rather less likely: 'adopt the majority view among the known views of one's friends'.

Combining the graph setting with epistemic update rules requires an enriched model-

ing in terms of agent graphs plus knowledge (cf. e.g., Baltag et al., 2019b; Christoff, 2016; Liu et al., 2014; Seligman et al., 2011), endowed with suitable update rules. In such a setting with knowledge made explicit, we can also ask quite new questions, such as whether position in the graph determines social influence. And also, a crucial phenomenon will be acts of *communication* where facts about neighbors come to be known.

Combining these ideas with our general approach, Chapter 5 provides a comprehensive proposal for a rich model of opinion formation in social groups via threshold models that includes modeling agents' knowledge explicitly. We study this setting and find the logic representing the basic reasoning about such social dynamical systems. This final chapter may be seen as a case study for what might be a systematic 'epistemization' of graph games and dependence logics.

This concludes our introduction to the main topics of the dissertation.

1.2 Outline of the thesis

Here is a brief summary of the concrete topics and results in this dissertation.

Chapter 2 looks at social scenarios in which a hostile agent may destroy the playground on which communication and interaction take place, as suggested by Example 1.1. To study such interactions with the techniques of graph games, we give a mathematical model where a player can remove links of the underlying graph with an explicit description of the targets to be blocked. Afterwards, we develop a modal logic of definable link deletion, which matches precisely with our games, in that its language is expressive enough to characterize the actions of players and determine their winning positions. Also, we settle a range of meta-properties of the resulting logic, using a new type of first-order translation for the logic. We also provide a notion of bisimulation that leads to a characterization theorem for the logic as a fragment of first-order logic, and allows us to compare the expressive power of our logic with that of known hybrid modal languages. Next, we discuss how to axiomatize this logic of link deletion, using dynamic-epistemic logics as a contrast. Finally, we show that our new modal logic lacks both the tree model property and the finite model property and that its satisfiability problem is undecidable.

In Chapter 3 we apply graph games with link deletion to social scenarios with positive goals. Motivated by Example 1.2, the chapter analyzes the interactions between agents in learning/teaching scenarios and proposes a comprehensive framework of 'learning games

with corrections', to capture realistic features of educational processes. A learner may make mistakes in the process, or ignore useful available information. On the other hand, a teacher can correct mistakes made by the learner, or highlight facts ignored by adding links to the graph. Based on such games, we provide a modal logic of correct learning whose models can represent correct and wrong inferences, with formulas evaluated at histories in the process of learning. We connect this richer dynamic language to existing modal and first-order logics, and again establish results on first-order translation, a matching notion of bisimulation and a characterization theorem. Additionally, we determine the computational complexity of the logic. In particular, the model checking problem is PSPACE-complete and the satisfiability problem is undecidable.

In Chapter 4 we turn to abstract dependence of behavior, with agents represented by variables, in dynamical systems. We first add a temporal dimension to the basic modal logic of functional dependence, in Baltag and van Benthem (2021b), and study the resulting system from both modal and first-order perspectives that, though equivalent, mutually complete each other. Again using results on bisimulation and translation, we chart the expressive power of the logic, and also, we present a complete Hilbert-style proof system based on a representation results for abstract dependence models as dynamical systems, and prove a decidability result for a significant fragment of the logic. Still in the same chapter, we also consider richer topological dynamical systems with continuous transition functions, and extend the analysis to a logical framework for 'dynamic continuous dependence', which is closer to actual practice in the empirical sciences. We identify several natural sorts of continuous dependence, and present a Hilbert-style calculus for the logic of one attractive proposal. This topological perspective brings together dependence logics and existing 'dynamic topological logics' in a way that suggests a range of new questions in the study of dynamical systems.

Moving a bit closer to social reality, Chapter 5 studies a significant embodiment of the abstract notion of dependence, namely, the diffusion of behaviors or opinions in communities. Motivated by (Baltag et al., 2019b; Christoff, 2016), we focus on threshold models that are commonly used in the social sciences. Here it turns out that, in order to characterize the phenomena well, an epistemic dimension must be incorporated in our models, to capture the fact that agents' reactions to others generally depend on information flow. Accordingly, our proposed logical system combines more standard update operators for opinion change with acts of communication and information exchange. Among a

number of technical results about this setting, we present several proof systems for opinion update logics with or without epistemic ingredients.

Finally, Chapter 6 summarizes the main results of the dissertation, and identifies a variety of further directions from more theoretical and more practical points of view.

1.3 Sources of the chapters

- Chapter 2 is based on:

  Dazhu Li (2020). Losing connection: the modal logic of definable link deletion. *Journal of Logic and Computation*, 30(3):715-743.

- Chapter 3 is based on:

  Alexandru Baltag, Dazhu Li, and Mina Young Pedersen (2019). On the right path: a modal logic for supervised learning. In *Proceedings of LORI 2019*, *Lecture Notes in Computer Science*, 2019, 11813:1-14. (Baltag et al., 2019c)

  Alexandru Baltag, Dazhu Li, and Mina Young Pedersen (2021). A modal logic for supervised learning. Accepted for publication in *Journal of Logic, Language and Information*.

- Chapter 4 is based on:

  Alexandru Baltag, Johan van Benthem, and Dazhu Li (2021). A logical analysis of dynamic dependence. *Manuscript*.

- Chapter 5 is based on:

  Alexandru Baltag, Dazhu Li, and Fernando R. Velàzquez-Quesada (2021). A logical approach to diffusion in social networks. *Manuscript*.

1.4 Technical preliminaries

In this part, we briefly provide a basic introduction to the technical preliminaries that are useful to understand the whole work of the dissertation.

1.4.1 Graph games and logics

Graph games are an important technical tool used to capture many social interactions in the dissertation. There are a number of those games having very tight relations with logics: sabotage games (van Benthem, 2014), poison games (Blando et al., 2020), Boolean

network games (Thompson, 2020) and correct learning games (Baltag et al., 2019c, 2021). We defer the overview of these works to Section 2.7. For now, since the works on sabotage games and its logic, the sabotage modal logic, play a significant role in the dissertation, we take them as an example to give the reader a basic feeling on the power of graph games to capture interactive scenarios and their connections with logics.

A sabotage game SG is played on a graph with a start-node and a goal-node. Also, there are two players *Traveler* and *Blocker*. In each round, Blocker deletes one link from the graph, and then, to arrive at the goal-node, Traveler moves along a link that is still available. Finally, the game is zero-sum: Traveler wins if she finally arrives at the goal, otherwise she loses. Here is an example.

Example 1.4: Consider the graph depicted in Figure 1.3, where a is the start-node and G is the goal-node. In this setting, who will win? In fact, it turns out that Blocker has a winning strategy. For instance, Blocker may start by deleting one of the links between node c and G, then Traveler moves to b. In the second round, Blocker has to remove the link between b and G, and Traveler now can move to c. Then, Blocker deletes the other link between c and G. Now, it is obvious that Traveler will get stuck finally.

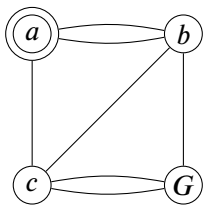


Figure 1.3 A graph.

In terms of games, how agents are allowed to act is important: different designs for their actions definitely give rise to different sorts of games. Here it is useful to briefly common on the actions of Blocker:

- Stepwise: only one link can be deleted at a time.
- Global: links deleted need not connect the current position of Traveler.

From a logical point of view, sabotage games are captured by the *sabotage modal logic* SML. Formally, its language $\mathcal{L}_{\blacklozenge}$ is defined as follows:

Definition 1.1: Let $\mathbf{P}$ be a countable set of propositional atoms. The *language* $\mathcal{L}_{\blacklozenge}$ *of the sabotage modal logic* is generated by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Diamond\varphi \mid \blacklozenge\varphi$$

where $p \in \mathbf{P}$. Also, define $\varphi \vee \psi := \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \to \psi := \neg\varphi \vee \psi$, $\varphi \leftrightarrow \psi := (\varphi \to \psi) \wedge (\psi \to \varphi)$, $\top := p \vee \neg p$, $\bot := \neg\top$, $\Box\varphi := \neg\Diamond\neg\varphi$ and $\blacksquare\varphi := \neg\blacklozenge\neg\varphi$.

Therefore, $\mathcal{L}_{\blacklozenge}$ enriches *the standard modal language* $\mathcal{L}_{\Box}$ with an additionally *sabotage modality* $\blacklozenge$. Intuitively, *the standard modality* $\Diamond$ characterizes the actions of Traveler in a SG, while $\blacklozenge$ is designed for that of Blocker. This will become clearer after we present their interpretations.

Formulas of $\mathcal{L}_{\blacklozenge}$ are evaluated in the *standard relational models* $\mathcal{M} = \langle W, R, V\rangle$, where W is a non-empty set of possible worlds or states, $R \subseteq W \times W$ is a binary relation and $V : \mathbf{P} \to \mathcal{P}(W)$ is a valuation function.[1] Also, a pair $\mathcal{F} = \langle W, R\rangle$ is called a *frame*. For each $w \in W$, $\langle \mathcal{M}, w\rangle$ is a *pointed model*. For brevity, we usually write $\mathcal{M}, w$ instead of $\langle \mathcal{M}, w\rangle$. We now introduce the semantics of SML, which is defined inductively by truth conditions.

Definition 1.2: Let $\mathcal{M} = \langle W, R, V\rangle$ be a model, $w \in W$ and $\varphi \in \mathcal{L}_{\blacklozenge}$. The *semantics of* SML is defined as follows:

$$
\begin{aligned}
\mathcal{M}, w \vDash p &\text{ iff } w \in V(p)\\
\mathcal{M}, w \vDash \neg\varphi &\text{ iff not } \mathcal{M}, w \vDash \varphi\\
\mathcal{M}, w \vDash \varphi \wedge \psi &\text{ iff } \mathcal{M}, w \vDash \varphi \text{ and } \mathcal{M}, w \vDash \psi\\
\mathcal{M}, w \vDash \Diamond\varphi &\text{ iff there exists } v \in W \text{ s.t. } Rwv \text{ and } \mathcal{M}, v \vDash \varphi\\
\mathcal{M}, w \vDash \blacklozenge\varphi &\text{ iff there are } s, t \in W \text{ s.t. } Rst \text{ and } \langle W, R \setminus \{\langle s, t\rangle\}, V\rangle, w \vDash \varphi
\end{aligned}
$$

From the perspective of SG, formula $\Diamond\varphi$ states that *the current position of Traveler has access to a point such that after she moves there* φ *holds*, while $\blacklozenge\varphi$ states that *there is a link such that after Blocker removes the link,* φ *is the case*. Also, formulas of $\mathcal{L}_{\blacklozenge}$ are able to describe winning positions for players in SG. For instance, that for Blocker in Example 1.4 is as follows:

$$\neg p \wedge \blacklozenge\Box(\neg p \wedge \blacklozenge\Box(\neg p \wedge \blacklozenge\Box(\neg p \wedge \blacklozenge\Box(\neg p \wedge \blacklozenge\Box(\neg p \wedge \blacklozenge\Box\bot))))))$$

where p is a distinguished propositional letter that is true only at the goal-node G. However, usually this pattern does not work on infinite games. To capture those for the latter, in general we need to extend SML with a modal μ-calculus, but we do not consider infinite games in the dissertation.

1 For any set A, $\mathcal{P}(A)$ refers to its power set.

For the logical aspect, augmenting $\mathcal{L}_{\Box}$ with the sabotage modality increases the expressive power drastically. For instance, different from the standard modal logic, SML is able to counter successors of the current point. But meanwhile, this also bring us some side effects. Say, the complexity of SML is high: its satisfiability problem and model checking problem have been proved to be undecidable and PSPACE-complete respectively. Also, validities in SML are not invariant under substitution. Furthermore, although the logic is axiomatizable, we lack a good Hilbert-style calculus with language $\mathcal{L}_{\blacklozenge}$.

1.4.2 Dynamic-epistemic logics

To understand the contents of the dissertation, it is useful to be familiar with some basics of dynamic-epistemic logics DEL (Baltag et al., 1998; Plaza, 1989; van Benthem, 2011; van Ditmarsch et al., 2007), which are mentioned throughout the dissertation. The family of the logical frameworks is famous for their success in modelling of and reasoning about dynamics of epistemic states of agents, induced by learning more about facts. This part restricts itself to one of the simplest yet best known frameworks within the family of DEL, i.e., *the public-announcement logic* PAL for single agent. Here is the formal language:

Definition 1.3: Let $\mathbf{P}$ be a countable set of propositional atoms. The *language* $\mathcal{L}_p$ *of public announcement logic* is defined as follows

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K\varphi \mid [!\varphi]\varphi$$

where $p \in \mathbf{P}$. Usually, we call K the *knowledge operator* and $[!\ ]$ the *announcement operator*.

Intuitively, formula $K\varphi$ states that *the agent knows* φ, and $[!\varphi]\psi$ expresses *if* φ *is announcable, then* ψ *is the case after* φ *is publicly announced*. Here the precondition for announcing a formula φ is that φ should be true at the actual state.

Models $\mathcal{M} = \langle W, \sim, V \rangle$ of PAL are the same as standard relational models, except that $\sim$ is an *equivalence relation*, i.e., it is reflexive, transitive and symmetric. Intuitively, the relation represents the 'indistinguishability' of different situations (from the perspective of the agent in the model). That is, for any $w, v \in W$, $w \sim v$ intuitively means that the agent cannot tell w from v. Formally, the semantics of PAL is as follows:

Definition 1.4: Let $\mathcal{M} = \langle W, \sim, V \rangle$ be a model of PAL, $w \in W$ and $\varphi \in \mathcal{L}_p$. The *truth conditions* for propositional atoms and Boolean connectives are the same as those in

Definition 1.2. Moreover,

$$\mathcal{M}, w \vDash K\varphi \text{ iff for all } v \in W, \text{ if } w \sim v, \text{ then } \mathcal{M}, v \vDash \varphi$$
$$\mathcal{M}, w \vDash [!\varphi]\psi \text{ iff if } \mathcal{M}, w \vDash \varphi \text{ then } \mathcal{M}^{\varphi}, w \vDash \psi$$

where $\mathcal{M}^{\varphi} = \langle W, \sim^{\varphi^{\mathcal{M}}}, V\rangle$ and $s \sim^{\varphi^{\mathcal{M}}} t$ iff (a). $s \sim t$ and (b). $\mathcal{M}, s \vDash \varphi \Leftrightarrow \mathcal{M}, t \vDash \varphi$.[1]

Intuitively, the clause for knowledge operator $K\varphi$ means that the agent is regarded to know φ if all situations considered by her are φ (i.e., she rules out all situations that are not φ from her consideration). It is important to recognize that the resulting $\sim^{\varphi^{\mathcal{M}}}$ is still an equivalence relation, i.e., the update is well-defined on models of PAL. Intuitively, the update indicates that: now the agent knows that φ was the case and all other situations can be distinguished from those cases. The update has the several salient features:

- Uniform: all links that do not satisfy the condition are removed simultaneously.
- Global: links deleted need not connect the actual world.

The PAL without public announcement operator is precisely captured by the well-known proof system **S5**. Perhaps surprisingly, although PAL has additional dynamic operators, its expressive power is the same as that of **S5**, which can illustrated by the following '*recursion axioms*' for operator [!]:

$$[!\varphi]p \leftrightarrow \varphi \rightarrow p$$
$$[!\varphi]\neg\psi \leftrightarrow \varphi \rightarrow \neg[!\varphi]\psi$$
$$[!\varphi](\psi \wedge \chi) \leftrightarrow [!\varphi]\psi \wedge [!\varphi]\chi$$
$$[!\varphi]K\psi \leftrightarrow \varphi \rightarrow K(\varphi \rightarrow [!\varphi]\psi)$$

The key spirit of the axioms is that all dynamic formulas can be recursively reduced to the static fragment of $\mathcal{L}_p$. The feasibility of doing this definitely depends heavily on the method of updates induced by the public announcement operator: even very slight changes to the features mentioned above may stop the recursive format. However, the collapse of the public announcement logic into **S5** by no means states that they are identical. A piece of evidence to illustrate this is that: the set of validities of **S5** is closed under substitution, while that of PAL is not.

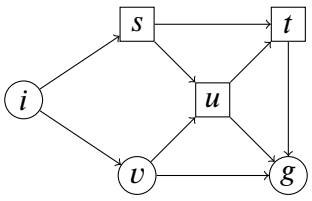
1 It might be more popular to define $\mathcal{M}^{\varphi} = \langle W', \sim', V'\rangle$ as restriction of $\mathcal{M}$ to the worlds where φ is true, i.e, $W' := \{w \in W \mid \mathcal{M}, w \vDash \varphi\}$, $\sim' := \sim \cap (W' \times W')$ and $V'(p) := V(p) \cap W'$. The basic results introduced in this part also apply to this kind of updates. But, to highlight the intrinsic difference between PAL and SML, it is useful to adopt the definition involving link-deletion.

Chapter 2 A logic for graph games with definable link deletions

2.1 Social interactions under adverse circumstances

Interactions between social agents take many forms, studied in the social sciences, mathematics, computer science, philosophy, and other fields. Not surprisingly, new perspectives and new formal models keep arising. In this chapter, we take a look at interactive scenarios where players may obstruct each other by changing the environment in which their interactions take place. This phenomenon, introduced with various examples in Chapter 1 is ubiquitous. For instance, in social networks, players may delete or add friends, and this clearly affects the environment in which they form or transmit their opinions or adjust their behaviors. We start with a mathematical framework that can model these scenarios in a precise way, namely, *graph games*.

Let us first recall Example 1.1 introduced in the previous chapter.



Now Alice starts at point i, and tries to arrive at one of the goal points t and g. However, Bob, the spy travelling with her, tries to prevent this.[1] The game goes in rounds: Bob first cuts one or more links in the graph, then Alice makes a step along one of the links still available. Since Bob can cut at most 9 links in all, the game is finite. Alice wins if she gets to one of the goal regions, and loses if she cannot get there.

This description still leaves the game underspecified, since we must say more about how Bob is allowed to cut before we can analyze the outcomes of the game. For concreteness, we start with a variant where the properties are not yet essential.

First version. Bob cuts one arrow from Alice's current position to some reachable node.

In the resulting game on our graph, Alice has a winning strategy: she is always able to arrives at one of the regions. Bob might start by deleting the link $\langle i, s\rangle$, then Alice moves

1 For simplicity, we ignore other members of the criminal group, and just consider the case that it is Bob who is destroying the roads.

to node v. In the second round, Bob must cut $\langle v, g\rangle$, and Alice goes to state u. Finally, player Alice can always arrive at t or g whatever link Bob deletes.

In this first version, the game is a local variant of the *sabotage game* SG (see, e.g., van Benthem, 2014). A sabotage game is played on a graph by two players: in each round, *Traveler* acts in the same way as Alice, while Bob's counterpart *Blocker* first cuts a link. However, *Blocker*'s moves in sabotage games are global and allow cutting a link anywhere in the graph, not necessarily starting at the current position of *Traveler*. In contrast, our game restricts the moves available to *Blocker*, giving him fewer winning strategies in general (cf. e.g., Areces et al., 2015; Aucher et al., 2018; Rohde, 2005).

However, the real-world scenario that we considered suggest a more drastic deviation from existing sabotage games. In particular, Bob in fact can destroy the roads with the properties of the locations that Alice likes. So, our next game models such more terrible scenario, taking care of both aspects.

Definitive version. In each round, player Bob chooses an available atomic property, and cuts all links from the position of Alice to nodes with the chosen property.

For example, in the graph depicted in Figure 1.1, when Alice is located at node s, Bob can cut both the links $\langle s, u\rangle$ and $\langle s, t\rangle$ if he chooses the definable property of nodes marked by the square.

Clearly, with this new version, Bob's powers of blocking access to information have increased. Indeed, on the same graph as before, he now has a winning strategy. In the first round, Bob cuts the link $\langle i, v\rangle$, and Alice's only option is to move to node s. But then, Bob can cut both links $\langle s, u\rangle$ and $\langle s, t\rangle$ simultaneously, and Alice gets stuck and loses.

We will now focus on the logical analysis of our second more realistic game, calling it the *definable sabotage game* $\mathsf{S_dG}$. Here existing modal logics for sabotage can serve as an inspiration, given the similarity of the games. But they must be modified, since we have made the obstructing player both less powerful (given the local nature of his choices) and more powerful (since he can remove more than one link in general). More concretely, to analyze the sabotage game, a *sabotage modal logic* SML is proposed, which extends standard modal logic with a sabotage modality $\blacklozenge\varphi$ stating that φ is true at the evaluation point after removing some accessibility arrow from the model (see, e.g., Aucher et al., 2018; Löding and Rohde, 2003a). But what is a suitable logic for $\mathsf{S_dG}$? The next section contains our proposal, called *definable sabotage modal logic* $\mathsf{S_dML}$. We will study this logic in depth, not just for its connections to the above games, but also as a pilot study

for throwing light on what is special and what is general about sabotage games, and the logical theory that already exists for them. In addition, our logic is a test case for how local sabotage, even though definable in ways reminiscent of dynamic-epistemic logics of information update, has its own behavior, including significantly higher complexity (cf. Areces et al., 2012, 2018).

Outline of the chapter. In Section 2.2, we present the syntax and semantics of $\mathsf{S_dML}$ (Section 2.2), and some typical logical validities (Section 2.2.2). In Section 2.3, we describe the non-trivial first-order translation for $\mathsf{S_dML}$ and check its correctness. In Section 2.4, we first introduce a notion of bisimulation for $\mathsf{S_dML}$ and investigate some of its model theory (Section 2.4.1), then we prove a characterization theorem for $\mathsf{S_dML}$ as a fragment of first-order logic that is invariant for the bisimulation introduced (Section 2.4.2), and finally we explore the expressive power of $\mathsf{S_dML}$ (Section 2.4.3). In Section 2.5, we provide some further analysis of an axiomatization of $\mathsf{S_dML}$. In particular, we illustrate the relation between $\mathsf{S_dML}$ and hybrid logics (Section 2.5.1), and study recursion axioms (Section 2.5.2). Next, in Section 2.6, we show that $\mathsf{S_dML}$ lacks both the tree model property and the finite model property, and that the satisfiability problem for $\mathsf{S_dML}$ is undecidable. Finally, we discuss related work in Section 2.7, and conclude in Section 2.8 with a summary and outlook on further directions.

2.2 The modal logic of $\mathsf{S_dG}$: $\mathsf{S_dML}$

In this section, we introduce the language and semantics of logic $\mathsf{S_dML}$. After that, to understand the new device, we illustrate some properties of the logic by means of logical validities.

2.2.1 Language and semantics

As mentioned above, the definable sabotage modal logic $\mathsf{S_dML}$ is intended to match $\mathsf{S_dG}$. Therefore its language should be expressive enough to model the actions of the players. For player Alice, it is natural to think of the standard modality $\Diamond$, which characterizes the transition from a node to its successors (see, e.g., Blackburn et al., 2001; van Benthem, 2010). However, to characterize the action of Bob, some dynamic operator is indispensable.

The language $\mathcal{L}_d$ of $\mathsf{S_dML}$ is a straightforward extension of the standard modal language $\mathcal{L}_\Box$. In addition to the modality $\Box$, it also includes a dynamic modal operator $[-]$.

The formal definition is as follows:

Definition 2.1: Let $\mathbf{P}$ be a countable set of propositional atoms. The *language* $\mathcal{L}_d$ is defined by the following grammar in Backus-Naur Form:

$$\mathcal{L}_d \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box\varphi \mid [-\varphi]\varphi$$

where $p \in \mathbf{P}$. Besides, notions $\top$, $\bot$, $\vee$, $\rightarrow$, $\leftrightarrow$ and $\Diamond$ are as usual. For any $[-\varphi]\psi \in \mathcal{L}_d$, we define $\langle -\varphi\rangle\psi := \neg[-\varphi]\neg\psi$, i.e., $\langle - \rangle$ is the dual operator of $[-]$.

We will often omit parentheses when doing so ought not to cause confusion. The operator $[-]$ is our device to model the action of Bob in $\mathsf{S_dG}$. This can be clarified by the semantics of $\mathsf{S_dML}$. Formulas of $\mathcal{L}_d$ are evaluated in standard relational models $\mathcal{M} = \langle W, R, V\rangle$. For any $\langle w, v\rangle \in R$, we also write $\langle w, v\rangle \in \mathcal{M}$. In addition, we use $R(w)$ to denote the set $\{v \in W \mid \langle w, v\rangle \in R\}$ of successors of w. We now introduce the semantics, which is defined inductively by truth conditions.

Definition 2.2: Let $\mathcal{M} = \langle W, R, V\rangle$ be a model, $w \in W$ and $\varphi \in \mathcal{L}_d$. The *semantics* for the language $\mathcal{L}_d$ is defined as follows:

$$\begin{aligned}
\mathcal{M}, w \vDash p &\text{ iff } w \in V(p)\\
\mathcal{M}, w \vDash \neg\varphi &\text{ iff } \mathcal{M}, w \nvDash \varphi\\
\mathcal{M}, w \vDash \varphi \wedge \psi &\text{ iff } \mathcal{M}, w \vDash \varphi \text{ and } \mathcal{M}, w \vDash \psi\\
\mathcal{M}, w \vDash \Box\varphi &\text{ iff for each } v \in W, \text{ if } Rwv, \text{ then } \mathcal{M}, v \vDash \varphi\\
\mathcal{M}, w \vDash [-\varphi]\psi &\text{ iff } \mathcal{M}|_{\langle w,\varphi\rangle}, w \vDash \psi
\end{aligned}$$

where $\mathcal{M}|_{\langle w,\varphi\rangle} = \langle W, R \setminus (\{w\} \times \{u \in R(w) \mid \mathcal{M}, u \vDash \varphi\}), V\rangle = \langle W, R \setminus (\{w\} \times \{u \in W \mid \mathcal{M}, u \vDash \varphi\}), V\rangle$ is obtained by deleting all links from w to the nodes that are φ.

We let $\|\varphi\|^{\mathcal{M}} = \{w \in W \mid \mathcal{M}, w \vDash \varphi\}$ denote the *truth set* of a formula φ in $\mathcal{M}$. We omit the superscript for the model when it is clear from the context. A formula φ is *satisfiable* if there exists a pointed model $\langle \mathcal{M}, w\rangle$ with $w \in \|\varphi\|$. By Definition 2.2, the truth conditions for Boolean and modal connectives $\neg$, $\wedge$, $\Box$ are as usual, and $[-\varphi]\psi$ states that ψ is true at the evaluation point after deleting all links from the current point to the nodes that are φ. Intuitively, by the semantics, formula φ occurring in $[-]$ stands for a property of some successors of the current point, and $[-\varphi]$ represents an action of Bob in $\mathsf{S_dG}$.

Example 2.1: Recall the driving map of Alice. Assume that the propositional atoms p and q refer to the properties denoted with circle and square respectively. Then we are able to express the facts of the game with formulas of $\mathcal{L}_d$. For instance, that 'after Bob deletes the links from v to the circle point, i.e., g, Alice still can move to a square node, i.e., u' can be expressed as the truth at v of the formula $[-p]\Diamond q$. Moreover, $\mathcal{L}_d$ can also describe the winning strategies for players in this example. Say, the formula $[-p]\Box[-q]\Box\bot$ states that Bob can stop Alice successfully by removing the links from the position of Alice to the circle nodes in the first round, and cutting the links pointing to the square nodes in the second round. By our semantics for these formulas, $\mathsf{S_dML}$ is suitable to capture $\mathsf{S_dG}$.

2.2.2 Logical validities

Although the language and semantics of $\mathsf{S_dML}$ look simple, there are some issues with the new operator $[-]$. To illustrate how it works, we explore some interesting validities of $\mathsf{S_dML}$. First of all, let us consider the following principle:

$$[-\varphi](\varphi_1 \to \varphi_2) \to ([-\varphi]\varphi_1 \to [-\varphi]\varphi_2) \tag{2-1}$$

which follows from the semantics of $\mathsf{S_dML}$ directly. The formula enables us to distribute $[-]$ over an implication. It is a common principle that applies to almost all modalities, such as the standard modality and the public announcement operator (see, e.g., Baltag et al., 1998). However, operator $[-]$ also has some distinguishing features. For instance, the validity

$$[-\varphi]\psi \leftrightarrow \langle -\varphi\rangle\psi \tag{2-2}$$

illustrates that $[-]$ is self-dual and—less obviously—a model update function essentially. It is not hard to check that the validity of formulas (2-1) and (2-2) is closed under substitution. Interestingly, this is not a common feature of $\mathsf{S_dML}$. Some examples are as follows:

$$[-\varphi]p \leftrightarrow p \tag{2-3}$$

$$[-p]\Diamond q \leftrightarrow \Diamond(\neg p \wedge q) \tag{2-4}$$

$$[-p][-q]\varphi \leftrightarrow [-q][-p]\varphi \tag{2-5}$$

Principle (2-3) illustrates that operator $[-]$ does not change the truth value of propositional atoms. Formula (2-4) allows us to reduce a formula including $[-]$ to an $\mathcal{L}_\Box$-formula. By

(2-5), when all formulas occurring in $[-]$ are propositional atoms, the order of different operators $[-]$ can be interchanged.

Actually each propositional atom occurring in formulas (2-3)-(2-5) can be replaced by any Boolean formula without affecting their validity. However, these schematic validities fail in general when we consider the deletions for complex properties. See Figure 2.1 for an example showing this phenomenon for principle (2-5).

Figure 2.1 A model showing that the general schema $[-\varphi_1][-\varphi_2]\varphi \leftrightarrow [-\varphi_2][-\varphi_1]\varphi$ of principle (2-5) fails in $\mathsf{S_dML}$. Let $\varphi_1 := p$, $\varphi_2 := \Diamond\Diamond p$ and $\varphi := \Diamond q$. Then at point w formula $[-p][-\Diamond\Diamond p]\Diamond q$ is true, while $[-\Diamond\Diamond p][-p]\Diamond q$ fails.

Many instances of validity in $\mathsf{S_dML}$ are not straightforward, and require much more thought than the often rather obvious validities found in standard logical systems. In particular, the dynamic modality $[-]$ creates interesting complexity, since removing links in a model can have side-effects for truth values of formulas at worlds throughout the model. Therefore, it is time to make a deeper technical investigation of our logic.

2.3 A first-order translation for $\mathsf{S_dML}$

Given the semantics of $\mathsf{S_dML}$, a natural question is: is $\mathsf{S_dML}$ axiomatizable? By the completeness theorem for first-order logic, validity of first-order logic is effectively axiomatizable. Therefore, a positive answer to the question can be provided if we can describe a recursive standard translation for $\mathsf{S_dML}$ (cf. van Benthem, 1984, 2010).

Obviously, truth conditions for $\mathsf{S_dML}$ are first-order. So, in there must be a first-order translation like that for standard modal logic. However we already know from SML that additional arguments may be needed in the translation: for SML, that extra argument is a finite set of links (see Areces et al., 2015; Aucher et al., 2015). Interestingly, finding the translation here requires even more delicate analysis of the extra argument.

To do so, our method is to introduce a new device, being a sequence consisting of ordered pairs (e.g., $\langle y, \varphi\rangle$), to denote the occurrences of $[-]$ in a formula, where y is a variable and φ is a property of its successors. Let $\mathcal{L}_1$ be the first-order language consisting of countable unary predicates $P_{i\in N}$, a binary relation R and equality $\equiv$.

Definition 2.3: Let O be a finite sequence $\langle y_0, \psi_0\rangle; \cdots; \langle y_i, \psi_i\rangle; \cdots; \langle y_n, \psi_n\rangle$ $(0 \leqslant i \leqslant n)$ such that $\psi_{0\leqslant i\leqslant n}$ is an $\mathcal{L}_d$-formula and $y_{0\leqslant i\leqslant n}$ is a variable, and x be a designated variable.

The *standard translation* $ST_x^O : \mathcal{L}_d \to \mathcal{L}_1$ is defined recursively as follows:

$$\begin{aligned}
ST_x^O(p) &= Px \\
ST_x^O(\top) &= x \equiv x \\
ST_x^O(\neg\varphi) &= \neg ST_x^O(\varphi) \\
ST_x^O(\varphi_1 \wedge \varphi_2) &= ST_x^O(\varphi_1) \wedge ST_x^O(\varphi_2) \\
ST_x^O(\Diamond\varphi) &= \exists y(Rxy \wedge \neg(x \equiv y_0 \wedge ST_y^{\langle x,\bot\rangle}(\psi_0)) \wedge \\
&\qquad \bigwedge_{0\leqslant i\leqslant n-1} \neg(x \equiv y_{i+1} \wedge ST_y^{\langle y_0,\psi_0\rangle;\cdots;\langle y_i,\psi_i\rangle}(\psi_{i+1})) \wedge ST_y^O(\varphi)) \\
ST_x^O([-\varphi_1]\varphi_2) &= ST_x^{O;\langle x,\varphi_1\rangle}(\varphi_2)
\end{aligned}$$

where y is a variable which has not been used yet in the translation.

The key inductive clauses in Definition 2.3 concern $\Diamond$-formulas and $[-\ ]$-formulas. Formula $\Diamond\varphi$ is translated as a first-order formula with x free, stating that the current point x has a successor y satisfying $ST_y^O(\varphi)$, and that this edge is not deleted by the operator $[-\ ]$ indexed in the sequence O. The first-order translation for $[-\varphi_1]\varphi_2$ says that the translation of φ_2 is carried out with respect to the sequence $O;\langle x,\varphi_1\rangle$, and that this translation is realized at the current point x.

According to Definition 2.3, the index sequence O may become longer and longer, but it is always finite. For each formula φ of $\mathcal{L}_d$, $ST_x^{\langle x,\bot\rangle}(\varphi)$ yields a first-order formula with only x free. Now we use an example to illustrate the translation.

Example 2.2: Consider formula $\Diamond[-\Diamond p_1]\Box p_2$. Its translation runs as follows:

$$\begin{aligned}
ST_x^{\langle x,\bot\rangle}(\Diamond[-\Diamond p_1]\Box p_2) &= \exists y(Rxy \wedge \neg(x \equiv x \wedge ST_y^{\langle x,\bot\rangle}(\bot)) \wedge \\
&\qquad ST_y^{\langle x,\bot\rangle}([-\Diamond p_1]\Box p_2)) \\
&= \exists y(Rxy \wedge \neg(x \equiv x \wedge ST_y^{\langle x,\bot\rangle}(\bot)) \wedge \\
&\quad ST_y^{\langle x,\bot\rangle;\langle y,\Diamond p_1\rangle}(\Box p_2)) \\
&= \exists y(Rxy \wedge \neg(x \equiv x \wedge ST_y^{\langle x,\bot\rangle}(\bot)) \wedge \\
&\quad \forall z(Ryz \wedge \neg(y \equiv x \wedge ST_z^{\langle x,\bot\rangle}(\bot)) \wedge \\
&\quad \neg(y \equiv y \wedge ST_z^{\langle x,\bot\rangle}(\Diamond p_1)) \to ST_z^{\langle x,\bot\rangle;\langle y,\Diamond p_1\rangle}(p_2)) \\
&= \exists y(Rxy \wedge \neg(x \equiv x \wedge ST_y^{\langle x,\bot\rangle}(\bot)) \wedge \\
&\quad \forall z(Ryz \wedge \neg(y \equiv x \wedge ST_z^{\langle x,\bot\rangle}(\bot)) \wedge
\end{aligned}$$

$$\begin{aligned}
&\neg(y \equiv y \wedge \exists z'(Rzz' \wedge \neg(z \equiv x \wedge ST_{z'}^{\langle x,\bot\rangle}(\bot)) \wedge \\
&ST_{z'}^{\langle x,\bot\rangle}(p_1)) \rightarrow ST_z^{\langle x,\bot\rangle;\langle y,\Diamond p_1\rangle}(p_2)) \\
=\ &\exists y(Rxy \wedge \neg(x \equiv x \wedge \neg y \equiv y) \wedge \forall z(Ryz \wedge \\
&\neg(y \equiv x \wedge \neg z \equiv z) \wedge \neg(y \equiv y \wedge \exists z'(Rzz' \wedge \\
&\neg(z \equiv x \wedge \neg z' \equiv z') \wedge P_1 z') \rightarrow P_2 z)
\end{aligned}$$

The resulting formula is very complicated. Essentially, it is equivalent to formula $\exists y(Rxy \wedge \forall z(Ryz \wedge \neg\exists z'(Rzz' \wedge P_1 z') \rightarrow P_2 z))$, which states that there exists a successor y of the current point x such that, for each successor z of y, if z does not have any P_1-successors, then z is P_2. Example 2.2 can be considered as a small case illustrating that $\mathsf{S_dML}$ is succinct notation for a complex part of first-order logic. In order to check the result, we will prove the correctness of Definition 2.3. In what follows, for any assignment σ, we define $\sigma_x^w(x) = w$, and $\sigma_x^w(y) = \sigma(y)$ when $x \neq y$. Now let us first introduce the following lemma:

Lemma 2.1: Let $O = \langle y_0, \psi_0\rangle; \cdots; \langle y_i, \psi_i\rangle; \cdots; \langle y_n, \psi_n\rangle$ be a finite sequence, y a variable not occurring in O, σ an assignment, and $\langle \mathcal{M}, w\rangle$ a pointed model. Assume that ψ is an $\mathcal{L}_d$-formula such that:

(a). $\mathcal{M}, w \vDash \psi$ iff $\mathcal{M} \vDash ST_x^{\langle x,\bot\rangle}(\psi)[\sigma_x^w]$.

For any variable z and $i \leqslant n-1$, suppose that:

(b). $\mathcal{M}|_{\langle u,\psi\rangle} \vDash ST_z^{\langle x,\bot\rangle}(\psi_0)[\sigma]$ iff $\mathcal{M} \vDash ST_z^{\langle y,\psi\rangle}(\psi_0)[\sigma_y^u]$; and

(c). $\mathcal{M}|_{\langle u,\psi\rangle} \vDash ST_z^{\langle y_0,\psi_0\rangle;\cdots;\langle y_i,\psi_i\rangle}(\psi_{i+1})[\sigma]$ iff $\mathcal{M} \vDash ST_z^{\langle y,\psi\rangle;\langle y_0,\psi_0\rangle;\cdots;\langle y_i,\psi_i\rangle}(\psi_{i+1})[\sigma_y^u]$.

Then, it holds that:

$$\mathcal{M}|_{\langle u,\psi\rangle} \vDash ST_x^O(\varphi)[\sigma] \Leftrightarrow \mathcal{M} \vDash ST_x^{\langle y,\psi\rangle;O}(\varphi)[\sigma_y^u].$$

Proof It can be proven by induction on φ. The Boolean cases are routine, and we now consider other cases.

(1). φ is $\Diamond\varphi_1$. Assume that $\mathcal{M}|_{\langle u,\psi\rangle} \vDash ST_x^O(\varphi)[\sigma]$. Then in $\mathcal{M}|_{\langle u,\psi\rangle}$ we have a link $\langle \sigma(x), v\rangle$ such that $ST_z^O(\varphi_1)$, $\neg(x \equiv y_0 \wedge ST_z^{\langle x,\bot\rangle}(\psi_0))$ and $\neg(x \equiv y_{i+1} \wedge ST_z^{\langle y_0,\psi_0\rangle;\cdots;\langle y_i,\psi_i\rangle}(\psi_{i+1}))$ for any $i \leqslant n-1$, where $\sigma(z) = v$. By the inductive hypothesis, $\mathcal{M} \vDash ST_z^{\langle y,\psi\rangle;O}(\varphi_1)[\sigma_y^u]$. Moreover, from assumption (b), we can obtain $\mathcal{M} \vDash \neg(x \equiv y_0 \wedge ST_z^{\langle y,\psi\rangle}(\psi_0))[\sigma_y^u]$. Next, by assumption (c), for any $i \leqslant n-1$, formula $\neg(x \equiv y_{i+1} \wedge ST_z^{\langle y,\psi\rangle;\langle y_0,\psi_0\rangle;\cdots;\langle y_i,\psi_i\rangle}(\psi_{i+1}))$ is satisfied in $\mathcal{M}$ (with the assignment

σ_y^u). Furthermore, since $\langle \sigma(x), v\rangle \in \mathcal{M}|_{\langle u,\psi\rangle}$, we have $\mathcal{M}, v \not\models \psi$ if $\sigma(x) = u$. By assumption (a), we have $\mathcal{M} \models \neg(x \equiv y \wedge ST_y^{\langle y,\bot\rangle}(\psi))[\sigma_y^u]$. By Definition 2.3, it holds directly that $\mathcal{M} \models ST_x^{\langle y,\psi\rangle;O}(\varphi)[\sigma_y^u]$. Conversely, it is not hard to check that $\mathcal{M} \models ST_x^{\langle y,\psi\rangle;O}(\varphi)[\sigma_y^u]$ is followed by $\mathcal{M}|_{\langle u,\psi\rangle} \models ST_x^O(\varphi)[\sigma]$.

(2). φ is $[-\varphi_1]\varphi_2$. Suppose that $\mathcal{M} \models ST_x^{\langle y,\psi\rangle;O}(\varphi)[\sigma_y^u]$. By Definition 2.3, we have $\mathcal{M} \models ST_x^{\langle y,\psi\rangle;O;\langle x,\varphi_1\rangle}(\varphi_2)[\sigma_y^u]$. By the inductive hypothesis, $\mathcal{M}|_{\langle u,\psi\rangle} \models ST_x^O(\varphi_1)[\sigma]$ iff $\mathcal{M} \models ST_x^{\langle y,\psi\rangle;O}(\varphi_1)[\sigma_y^u]$. Therefore, the new sequence $O;\langle x,\varphi_1\rangle$ satisfies the assumption (c). Again, by the inductive hypothesis, $\mathcal{M}|_{\langle u,\psi\rangle} \models ST_x^{O;\langle x,\varphi_1\rangle}(\varphi_2)[\sigma]$. From Definition 2.3, we know $\mathcal{M}|_{\langle u,\psi\rangle} \models ST_x^O(\varphi)[\sigma]$. When $\mathcal{M}|_{\langle u,\psi\rangle} \models ST_x^O(\varphi)[\sigma]$, by an analogous argument in the converse direction, we can show $\mathcal{M} \models ST_x^{\langle y,\psi\rangle;O}(\varphi)[\sigma_y^u]$. ∎

With Lemma 2.1, we now are able to prove the correctness of the standard translation:

Theorem 2.1: Let σ be an assignment and $\varphi \in \mathcal{L}_d$. For any pointed model $\langle \mathcal{M}, w\rangle$, we have:

$$\mathcal{M}, w \models \varphi \Leftrightarrow \mathcal{M} \models ST_x^{\langle x,\bot\rangle}(\varphi)[\sigma_x^w].$$

Proof The proof is by induction on the structure of φ. The cases for Boolean and modal connectives are straightforward. When φ is $[-\varphi_1]\varphi_2$, the following equivalences hold:

$$\begin{aligned}
\mathcal{M}, w \models [-\varphi_1]\varphi_2 \ \text{ iff } \ & \mathcal{M}|_{\langle w,\varphi_1\rangle}, w \models \varphi_2 \\
\text{iff } \ & \mathcal{M}|_{\langle w,\varphi_1\rangle} \models ST_x^{\langle x,\bot\rangle}(\varphi_2)[\sigma_x^w] \\
\text{iff } \ & \mathcal{M} \models ST_x^{\langle x,\varphi_1\rangle;\langle x,\bot\rangle}(\varphi_2)[\sigma_x^w] \\
\text{iff } \ & \mathcal{M} \models ST_x^{\langle x,\bot\rangle;\langle x,\varphi_1\rangle}(\varphi_2)[\sigma_x^w] \\
\text{iff } \ & \mathcal{M} \models ST_x^{\langle x,\bot\rangle}(\varphi)[\sigma_x^w]
\end{aligned}$$

The first equivalence follows from the semantics directly. By the inductive hypothesis, for any pointed model $\langle \mathcal{M}_1, u\rangle$, we have $\mathcal{M}_1, u \models \varphi_2$ iff $\mathcal{M}_1 \models ST_x^{\langle x,\bot\rangle}(\varphi_2)[\sigma_x^u]$, therefore the second equivalence holds. Again, by the inductive hypothesis, we obtain $\mathcal{M}, w \models \varphi_1$ iff $\mathcal{M} \models ST_x^{\langle x,\bot\rangle}(\varphi_1)[\sigma_x^w]$. Therefore, the assumption (a) in Lemma 2.1 is satisfied. Besides, it is not hard to see that assumptions (b) and (c) in the lemma are also satisfied here. So, by Lemma 2.1, the third equivalence holds. Since no variable can satisfy the translation of $\bot$, the fourth one holds. The last one holds directly by Definition 2.3. ∎

Remark 2.1: The first-order translation for $\mathsf{S_dML}$ is quite different from that for sabotage modal logic. To translate a formula of sabotage modal logic, it suffices to maintain a

finite set of ordered pairs of nodes encoding the links already deleted. However it fails for $\mathsf{S_dML}$, since the number of links cut by $[-]$ may be infinite. In addition, we should also take care of the order of $[-]$ in a formula (recall Figure 2.1). Our finite sequence of ordered pairs of nodes and properties solves these problems and yields a suitable translation for $\mathsf{S_dML}$.

Finally, we end by answering the question stated at the outset of this section, which follows directly from Definition 2.3 and Theorem 2.1:

Corollary 2.1: From the completeness theorem for first-order logic, it follows that logic $\mathsf{S_dML}$ is axiomatizable.

2.4 Bisimulation and expressivity for $\mathsf{S_dML}$

Through the standard translation, we can translate a formula of $\mathsf{S_dML}$ into first-order logic syntactically. In this section, we investigate the other aspect, i.e., model theory, for its expressive power. Let us begin with considering the notion of bisimulation for $\mathsf{S_dML}$.

2.4.1 Bisimulation for $\mathsf{S_dML}$

After expanding the standard modal language $\mathcal{L}_\Box$ with the operator $[-]$, formulas of $\mathcal{L}_d$ are not invariant under the standard bisimulation (Blackburn et al., 2001) any longer.

To show this, let us first introduce a notion of *definable sabotage modal equivalence* (notation, $\leftrightsquigarrow_d$) between pointed models: $\langle \mathcal{M}_1, w\rangle \leftrightsquigarrow_d \langle \mathcal{M}_2, v\rangle$ iff for each $\varphi \in \mathcal{L}_d$, $\mathcal{M}_1, w \vDash \varphi$ iff $\mathcal{M}_2, v \vDash \varphi$.

Proposition 2.1: Formulas of $\mathcal{L}_d$ are not invariant under the standard bisimulation.

Proof It suffices to give an example. Consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$ as depicted in Figure 2.2. By the definition of the standard bisimulation, we know that both $\langle \mathcal{M}_1, w_1\rangle$ and $\langle \mathcal{M}_1, w_2\rangle$ are bisimilar to $\langle \mathcal{M}_2, v_1\rangle$, and that $\langle \mathcal{M}_1, w_3\rangle$ is bisimilar to $\langle \mathcal{M}_2, v_2\rangle$. However, we have $\mathcal{M}_1, w_1 \vDash [-q]\Diamond\Diamond q$ and $\mathcal{M}_2, v_1 \nvDash [-q]\Diamond\Diamond q$. Therefore, it follows that bisimulation does not imply definable sabotage modal equivalence. ∎

So, what is a suitable notion of bisimulation for $\mathsf{S_dML}$? Before answering this question, we first introduce some auxiliary definitions.

Definition 2.4: For any model $\mathcal{M} = \{W, R, V\}$ and $w \in W$, we say a set $U \subseteq W$ of possible worlds is *definable relative to* $R(w)$ in $\mathcal{M}$ iff there exists an $\mathcal{L}_d$-formula φ with

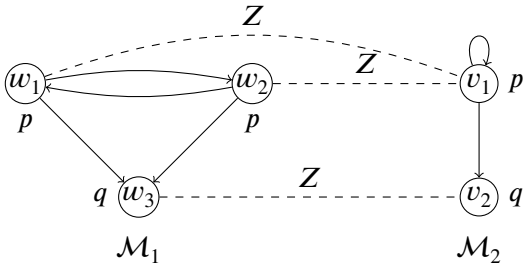


Figure 2.2 Two bisimilar models $\mathcal{M}_1$ and $\mathcal{M}_2$ (the bisimulation runs via the dashed lines labelled with 'Z').

$U = R(w) \cap \|\varphi\|$. Also, define $\mathcal{M}|_{\langle w,U\rangle} := \langle W, R \setminus (\{w\} \times U), V\rangle$ that is obtained by removing links $\{w\} \times U$ from model $\mathcal{M}$.

Take any formula φ of $\mathcal{L}_d$. It is not hard to see that the model $\mathcal{M}|_{\langle w,\|\varphi\|\rangle}$ is identical to $\mathcal{M}|_{\langle w,\varphi\rangle}$. Now we introduce a new notion of bisimulation for our logic $\mathsf{S_dML}$.

Definition 2.5: Let $\mathcal{M}_1 = \langle W_1, R_1, V_1\rangle$ and $\mathcal{M}_2 = \langle W_2, R_2, V_2\rangle$ be two models. We say a non-empty relation Z_d is a *definable sabotage bisimulation (d-bisimulation)* between pointed models $\langle \mathcal{M}_1, w\rangle$ and $\langle \mathcal{M}_2, v\rangle$ (notation, $\langle \mathcal{M}_1, w\rangle Z_d \langle \mathcal{M}_2, v\rangle$) if the following conditions are satisfied:

Atom: If $\langle \mathcal{M}_1, w\rangle Z_d \langle \mathcal{M}_2, v\rangle$, then $\mathcal{M}_1, w \vDash p$ iff $\mathcal{M}_2, v \vDash p$, for each $p \in \mathbf{P}$,

Zig$_\Diamond$: If $\langle \mathcal{M}_1, w\rangle Z_d \langle \mathcal{M}_2, v\rangle$ and there exists $w' \in W_1$ such that $R_1 w w'$, then there exists $v' \in W_2$ such that $R_2 v v'$ and $\langle \mathcal{M}_1, w'\rangle Z_d \langle \mathcal{M}_2, v'\rangle$,

Zag$_\Diamond$: If $\langle \mathcal{M}_1, w\rangle Z_d \langle \mathcal{M}_2, v\rangle$ and there exists $v' \in W_2$ such that $R_2 v v'$, then there exists $w' \in W_1$ such that $R_1 w w'$ and $\langle \mathcal{M}_1, w'\rangle Z_d \langle \mathcal{M}_2, v'\rangle$,

Zig$_{[-]}$: If $\langle \mathcal{M}_1, w\rangle Z_d \langle \mathcal{M}_2, v\rangle$ and U is definable relative to $R_1(w)$ in $\mathcal{M}_1$, then it holds that $\langle \mathcal{M}_1|_{\langle w,U\rangle}, w\rangle Z_d \langle \mathcal{M}_2|_{\langle v,Z_d(U)\rangle}, v\rangle$,

Zag$_{[-]}$: If $\langle \mathcal{M}_1, w\rangle Z_d \langle \mathcal{M}_2, v\rangle$ and U' is definable relative to $R_2(v)$ in $\mathcal{M}_2$, then it holds that $\langle \mathcal{M}_1|_{\langle w,Z_d^{-1}(U')\rangle}, w\rangle Z_d \langle \mathcal{M}_2|_{\langle v,U'\rangle}, v\rangle$.

where $Z_d(U) := \{v' \in R_2(v) \mid \langle \mathcal{M}_1, w'\rangle Z_d \langle \mathcal{M}_2, v'\rangle \text{ for some } w' \in U\}$, and $Z_d^{-1}(U') := \{w' \in R_1(w) \mid \langle \mathcal{M}_1, w'\rangle Z_d \langle \mathcal{M}_2, v'\rangle \text{ for some } v' \in U'\}$. We also write $\langle \mathcal{M}_1, w\rangle \underline{\leftrightarrow}_d \langle \mathcal{M}_2, v\rangle$ if there exists a d-bisimulation Z_d such that $\langle \mathcal{M}_1, w\rangle Z_d \langle \mathcal{M}_2, v\rangle$.

Here the conditions for $\Diamond$ are as usual, which do not change the model but change the evaluation point along the accessibility relation. In contrast, those for $[-]$ keep the evaluation point fixed but remove some links from the model. Besides, in standard modal logic, given any family of bisimulations $\{Z_i\}_{i\in I}$ between two models $\mathcal{M}$ and $\mathcal{N}$, the set-theoretic union $\bigcup\{Z_i\}_{i\in I}$ is again a bisimulation (see van Benthem, 2010). By Definition

2.5, this also holds for the new notion: the union of any family of d-bisimulations between two models is also a d-bisimulation. This observation is useful in various aspects, say, it can help us to simplify a given model to a smaller equivalent one.

Remark 2.2: It is worth noting that the clauses for our dynamic operator in Definition 2.5 are quite different from those defined by Areces et al. (2012); Fervari (2014) for the standard sabotage operator or relevant modalities. In particular, a special kind of definable sets are used. This is in line with the truth condition for $[-]$: we need to consider the properties of successors when updating the model at the current point. However, the usage of those definable sets also makes our notion of d-bisimulation intricate, since it essentially involves universally quantifying over all $\mathcal{L}_d$-formulas (recall Definition 2.4). That is, the notion defined in Definition 2.5 is syntax-dependent and thus not purely structural. It is a natural and interesting open problem whether this notion can be replaced by a fully structural one.[1]

As a concrete illustration of the d-bisimulation introduced here, it is easy to see that the pointed models $\langle \mathcal{M}_1, w_1 \rangle$ and $\langle \mathcal{M}_2, v_1 \rangle$ in Figure 2.2 are not d-bisimilar. Next we show that formulas of $\mathsf{S_dML}$ are invariant for d-bisimulation:

Theorem 2.2: For any pointed models $\langle \mathcal{M}_1, w \rangle$ and $\langle \mathcal{M}_2, v \rangle$, if $\langle \mathcal{M}_1, w \rangle \underline{\leftrightarrow}_d \langle \mathcal{M}_2, v \rangle$, then $\langle \mathcal{M}_1, w \rangle \nleftrightarrow_d \langle \mathcal{M}_2, v \rangle$.

Proof We prove it by induction on the syntax of φ. Let $\langle \mathcal{M}_1, w \rangle \underline{\leftrightarrow}_d \langle \mathcal{M}_2, v \rangle$.

(1). $\varphi \in \mathbf{P}$. By Definition 2.5, it holds directly that $\mathcal{M}_1, w \vDash \varphi$ iff $\mathcal{M}_2, v \vDash \varphi$.

(2). φ is $\neg\psi$. By the inductive hypothesis, $\mathcal{M}_1, w \vDash \psi$ iff $\mathcal{M}_2, v \vDash \psi$. Consequently, we know that $\mathcal{M}_1, w \vDash \varphi$ iff $\mathcal{M}_2, v \vDash \varphi$.

(3). φ is $\varphi_1 \wedge \varphi_2$. By the inductive hypothesis, for each $i \in \{1, 2\}$, $\mathcal{M}_1, w \vDash \varphi_i$ iff $\mathcal{M}_2, v \vDash \varphi_i$. Thus it holds that $\mathcal{M}_1, w \vDash \varphi$ iff $\mathcal{M}_2, v \vDash \varphi$.

(4). φ is $\Diamond\psi$. If $\mathcal{M}_1, w \vDash \varphi$, then there exists $w_1 \in W_1$ such that $R_1 w w_1$ and $\mathcal{M}_1, w_1 \vDash \psi$. By $\mathbf{Zig}_\Diamond$, there exists $v_1 \in W_2$ s.t. $R_2 v v_1$ and $\langle \mathcal{M}_1, w_1 \rangle \underline{\leftrightarrow}_d \langle \mathcal{M}_2, v_1 \rangle$. By

1 Instead of answering it, we refer the reader interested in this question to (Baltag and Cinà, 2018; Demey, 2011) which may be useful to solve this problem. Similar to our case, a non-structural notion of bisimulation for conditional belief on epistemic plausibility models is given by Demey (2011), which also discusses different methods, mainly by enhancing the logic or putting some special restrictions on models, to optimize the notion. Different from those methods used by Demey (2011), with the help of a notion of selection function, Baltag and Cinà (2018) provide a solid notion of bisimulation for conditional belief, behaving as desired both on plausibility models and on evidence models.

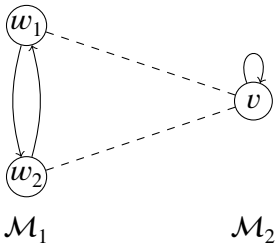


Figure 2.3 Two d-bisimilar models (the d-bisimulation runs through the dashed lines).

the inductive hypothesis, $\mathcal{M}_1, w_1 \vDash \psi$ iff $\mathcal{M}_2, v_1 \vDash \psi$. It is followed by $\mathcal{M}_2, v_1 \vDash \psi$ immediately. Consequently it holds that $\mathcal{M}_2, v \vDash \varphi$. Similarly, we can obtain $\mathcal{M}_1, w \vDash \varphi$ from $\mathcal{M}_2, v \vDash \varphi$ by $\mathbf{Zag}_{\Diamond}$.

(5). φ is $[-\varphi_1]\varphi_2$. Assume that $\mathcal{M}_1, w \vDash \varphi$. Define $U = \|\varphi_1\|^{\mathcal{M}_1} \cap R_1(w)$. Clearly, this set is definable relative to $R_1(w)$ in $\mathcal{M}_1$. By the truth condition for $[-\ ]$, we have $\mathcal{M}_1|_{\langle w,U\rangle}, w \vDash \varphi_2$. From $\mathbf{Zig}_{[-\ ]}$, we know that $\langle \mathcal{M}_1|_{\langle w,U\rangle}, w\rangle$ is d-bisimilar to $\langle \mathcal{M}_2|_{\langle v,Z_d(U)\rangle}, v\rangle$. By the inductive hypothesis, we have $\mathcal{M}_2|_{\langle v,Z_d(U)\rangle}, v \vDash \varphi_2$. To show $\mathcal{M}_2, v \vDash \varphi$, we now prove that for any $v' \in R_2(v)$, $v' \in \|\varphi_1\|^{\mathcal{M}_2}$ iff $v' \in Z_d(U)$.

Let $v' \in R_2(v)$. Suppose that $v' \in Z_d(U)$. Then, there is a $w' \in U$ such that $\langle \mathcal{M}_1, w'\rangle Z_d \langle \mathcal{M}_2, v'\rangle$. By the inductive hypothesis, we obtain $v' \in \|\varphi_1\|^{\mathcal{M}_2}$. For the other direction, let us assume that $v' \in \|\varphi_1\|^{\mathcal{M}_2}$. Since $R_2 v v'$, by $\mathbf{Zag}_{\Diamond}$ we know that w has a successor w' with $\langle \mathcal{M}_1, w'\rangle Z_d \langle \mathcal{M}_2, v'\rangle$. Consequently, by the inductive hypothesis, $\mathcal{M}_1, w' \vDash \varphi_1$. So, $w' \in U$. From the definition of $Z_d(U)$, we have $v' \in Z_d(U)$.

Therefore, $\mathcal{M}_2|_{\langle v,Z_d(U)\rangle}$ is identical to $\mathcal{M}_2|_{\langle v,\varphi_1\rangle}$. Consequently, $\mathcal{M}_2, v \vDash \varphi$. Similarly, by $\mathbf{Zag}_{[-\ ]}$, $\mathcal{M}_1, w \vDash \varphi$ follows from $\mathcal{M}_2, v \vDash \varphi$. ∎

As an application of Theorem 2.2, let us consider a simple example:

Example 2.3: Consider the models $\mathcal{M}_1$ and $\mathcal{M}_2$ depicted in Figure 2.3. By Definition 2.5, it holds that $\langle \mathcal{M}_1, w_1\rangle \underline{\leftrightarrow}_d \langle \mathcal{M}_2, v\rangle$ and $\langle \mathcal{M}_1, w_2\rangle \underline{\leftrightarrow}_d \langle \mathcal{M}_2, v\rangle$. From Theorem 2.2, we know that $\langle \mathcal{M}_1, w_1\rangle \nleftrightarrow_d \langle \mathcal{M}_2, v\rangle$ and $\langle \mathcal{M}_1, w_2\rangle \nleftrightarrow_d \langle \mathcal{M}_2, v\rangle$. Therefore, $\mathsf{S_dML}$ cannot distinguish between nodes $w_{1(2)}$ and v.

Furthermore, we can also show a weaker result for the other direction: for ω-saturated models, the converse of Theorem 2.2 holds as well. For each finite set Y, we denote the expansion of $\mathcal{L}_1$ with a set Y of constants with $\mathcal{L}_1^Y$, and denote the expansion of $\mathcal{M}$ to $\mathcal{L}_1^Y$ with $\mathcal{M}^Y$.

Definition 2.6: A model $\mathcal{M} = \langle W, R, V\rangle$ is *ω-saturated* if, for every finite subset Y of W, the expansion $\mathcal{M}^Y$ realizes every set $\Gamma(x)$ of $\mathcal{L}_1^Y$-formulas whose finite subsets $\Gamma'(x)$ are all realized in $\mathcal{M}^Y$.

From a standard modal point of view, Definition 2.6 requires that the evaluation point has some successors satisfying the set Γ of formulas if for any finite subset of Γ there are accessible states satisfying it. Besides, in terms of operator $[-]$, this definition requires that after an update at the evaluation state, if every finite subset of Γ is satisfied by the evaluation point, then the whole of Γ is satisfied by the point in the new model.

Not all models are ω-saturated, but every model can be extended to an ω-saturated model with the same first-order theory (see, e.g., Chang and Keisler, 1973). From Definition 2.3, we know that each model $\mathcal{M}$ has an ω-saturated extension with the same theory of $\mathsf{S_dML}$. For brevity, we use the set $\mathbb{T}^d(\mathcal{M}, w) = \{\varphi \in \mathcal{L}_d \mid \mathcal{M}, w \vDash \varphi\}$ to denote the $\mathsf{S_dML}$ **theory** of w in $\mathcal{M}$. With Definition 2.6, we end this part by the following result:

Theorem 2.3: For any ω-saturated $\langle \mathcal{M}_1, w\rangle$ and $\langle \mathcal{M}_2, v\rangle$, if $\langle \mathcal{M}_1, w\rangle \leftrightsquigarrow_d \langle \mathcal{M}_2, v\rangle$, then $\langle \mathcal{M}_1, w\rangle \underline{\leftrightarrow}_d \langle \mathcal{M}_2, v\rangle$.

Proof We prove this by showing that $\leftrightsquigarrow_d$ satisfies the definition of d-bisimulation.

(1). For each $p \in \mathbf{P}$, by the definition of $\leftrightsquigarrow_d$, it holds that $\mathcal{M}_1, w \vDash p$ iff $\mathcal{M}_2, v \vDash p$. This satisfies the condition of **Atom**.

(2). Let $w_1 \in W_1$ such that $R_1 w w_1$. We show that point v has a successor v_1 with $\langle \mathcal{M}_1, w_1\rangle \leftrightsquigarrow_d \langle \mathcal{M}_2, v_1\rangle$. For each finite subset Γ of $\mathbb{T}^d(\mathcal{M}_1, w_1)$, it holds that:

$$\begin{aligned} \mathcal{M}_1, w \vDash \Diamond \bigwedge \Gamma \ \text{ iff } \ & \mathcal{M}_2, v \vDash \Diamond \bigwedge \Gamma \\ \text{iff } \ & \mathcal{M}_2 \vDash ST_x^{\langle x,\bot\rangle}(\Diamond \bigwedge \Gamma)[\sigma_x^v] \\ \text{iff } \ & \mathcal{M}_2 \vDash \exists y(Rxy \wedge ST_y^{\langle x,\bot\rangle}(\bigwedge \Gamma))[\sigma_x^v] \end{aligned}$$

Therefore every finite subset Γ of $\mathbb{T}^d(\mathcal{M}_1, w_1)$ is satisfiable in the set of successors of node v. From Definition 2.6, we know that v has a successor v_1 where $\mathbb{T}^d(\mathcal{M}_1, w_1)$ is true. Thus, $\langle \mathcal{M}_1, w_1\rangle \leftrightsquigarrow_d \langle \mathcal{M}_2, v_1\rangle$. The proof of the $\mathbf{Zig}_\Diamond$ clause is completed.

(3). Similar to (2), we can prove that the condition of $\mathbf{Zag}_\Diamond$ is satisfied.

(4). Take any definable set $U = \|\varphi\|^{\mathcal{M}_1} \cap R_1(w)$ relative to $R_1(w)$ in $\mathcal{M}_1$. We prove $\mathbf{Zig}_{[-]}$ by showing $\langle \mathcal{M}_1|_{\langle w,U\rangle}, w\rangle \leftrightsquigarrow_d \langle \mathcal{M}_2|_{\langle v,Z_d(U)\rangle}, v\rangle$. For each finite subset Γ of $\mathbb{T}^d(\mathcal{M}_1|_{\langle w,U\rangle}, w)$, the following sequence of equivalences holds:

$$\mathcal{M}_1, w \vDash [-\varphi] \bigwedge \Gamma \ \text{ iff } \ \mathcal{M}_2, v \vDash [-\varphi] \bigwedge \Gamma$$

$$\text{iff } \mathcal{M}_2|_{\langle v,\varphi\rangle}, v \vDash \bigwedge \Gamma$$
$$\text{iff } \mathcal{M}_2|_{\langle v,\varphi\rangle} \vDash ST_x^{\langle x,\bot\rangle}(\bigwedge \Gamma)[\sigma_x^v]$$

From the proof of Theorem 2.2, we know that $\mathcal{M}_2|_{\langle v,\varphi\rangle}$ is exactly $\mathcal{M}_2|_{\langle v,Z_d(U)\rangle}$. So, it follows that each finite subset of $\mathbb{T}^d(\mathcal{M}_1|_{\langle w,U\rangle}, w)$ is true at v in $\mathcal{M}_2|_{\langle v,Z_d(U)\rangle}$. Then, by Definition 2.6, the theory $\mathbb{T}^d(\mathcal{M}_1|_{\langle w,U\rangle}, w)$ is true at v in $\mathcal{M}_2|_{\langle v,Z_d(U)\rangle}$. It is followed directly by $\langle \mathcal{M}_1|_{\langle w,U\rangle}, w\rangle \leftrightsquigarrow_d \langle \mathcal{M}_2|_{\langle v,Z_d(U)\rangle}, v\rangle$.

(5). Similar to (4), we can show that the condition of $\mathbf{Zag}_{[-]}$ is satisfied.

Thus, we conclude that $\langle \mathcal{M}_1, w\rangle \underline{\leftrightarrow}_d \langle \mathcal{M}_2, v\rangle$. The proof is completed. ∎

2.4.2 Characterization of $\mathsf{S_dML}$

With the notion of d-bisimulation, we can characterize $\mathsf{S_dML}$ as the one-free-variable fragment of first-order logic that is invariant for d-bisimulation, where a first-order formula $\alpha(x)$ is invariant for d-bisimulation just in case that for all pointed models $\langle \mathcal{M}_1, w_1\rangle$ and $\langle \mathcal{M}_2, w_2\rangle$ such that $\langle \mathcal{M}_1, w_1\rangle \underline{\leftrightarrow}_d \langle \mathcal{M}_2, w_2\rangle$, $\mathcal{M}_1 \vDash \alpha(x)[\sigma_x^{w_1}]$ iff $\mathcal{M}_2 \vDash \alpha(x)[\sigma_x^{w_2}]$.

Theorem 2.4: An $\mathcal{L}_1$-formula is equivalent to the translation of an $\mathcal{L}_d$-formula iff it is invariant for d-bisimulation.

Proof The direction from left to right holds directly by Theorem 2.2. For the converse direction, let α be an $\mathcal{L}_1$-formula with one free variable x. Assume that α is invariant for d-bisimulation. Now we consider the following set:

$$\mathbb{C}_d(\alpha) = \{ST_x^{\langle x,\bot\rangle}(\varphi) \mid \varphi \in \mathcal{L}_d \text{ and } \alpha \vDash ST_x^{\langle x,\bot\rangle}(\varphi)\}.$$

The result holds from the following two claims:

(i). If $\mathbb{C}_d(\alpha) \vDash \alpha$, then α is equivalent to the translation of an $\mathcal{L}_d$-formula.

(ii). $\mathbb{C}_d(\alpha) \vDash \alpha$, i.e., for any $\langle \mathcal{M}, w\rangle$, $\mathcal{M} \vDash \mathbb{C}_d(\alpha)[\sigma_x^w]$ entails $\mathcal{M} \vDash \alpha[\sigma_x^w]$.

We show (i) first. Suppose that $\mathbb{C}_d(\alpha) \vDash \alpha$. From the compactness and deduction theorems of first-order logic, it holds that $\vDash \bigwedge \Gamma \to \alpha$ for some finite subset Γ of $\mathbb{C}_d(\alpha)$. The converse can be shown by the definition of $\mathbb{C}_d(\alpha)$: $\vDash \alpha \to \bigwedge \Gamma$. Thus it holds that $\vDash \alpha \leftrightarrow \bigwedge \Gamma$ proving the claim.

As to the claim (ii), let $\langle \mathcal{M}, w\rangle$ be a pointed model such that $\mathcal{M} \vDash \mathbb{C}_d(\alpha)[\sigma_x^w]$. Consider the set $\Sigma = ST_x^{\langle x,\bot\rangle}(\mathbb{T}^d(\mathcal{M}, w)) \cup \{\alpha\}$. We now show that:

(a). The set Σ is consistent.

(b). $\mathcal{M} \vDash \alpha[\sigma_x^w]$, thus proving claim (ii).

Suppose that Σ is not consistent. By the compactness of first-order logic, it follows that $\models \alpha \to \neg \bigwedge \Gamma$ for some finite subset Γ of Σ. But then, by the definition of $\mathbb{C}_d(\alpha)$, we obtain $\neg \bigwedge \Gamma \in \mathbb{C}_d(\alpha)$, which is followed by $\neg \bigwedge \Gamma \in ST_x^{\langle x,\perp\rangle}(\mathbb{T}^d(\mathcal{M}, w))$. However, it contradicts to $\Gamma \subseteq ST_x^{\langle x,\perp\rangle}(\mathbb{T}^d(\mathcal{M}, w))$. Hence (a) holds.

Now we show that (b) holds as well. Since Σ is consistent, it can be realized by some pointed model, say, $\langle \mathcal{M}', w' \rangle$. Note that both the pointed models have the same $\mathsf{S_dML}$ theory, thus $\langle \mathcal{M}, w \rangle \leftrightsquigarrow_d \langle \mathcal{M}', w' \rangle$. Now take two ω-saturated elementary extensions $\langle \mathcal{M}_\omega, w \rangle$ and $\langle \mathcal{M}'_\omega, w' \rangle$ of $\langle \mathcal{M}, w \rangle$ and $\langle \mathcal{M}', w' \rangle$ respectively. It can be shown that such extensions always exist (see, e.g., Chang and Keisler, 1973). By the invariance of first-order logic under elementary extensions, from $\mathcal{M}' \models \alpha[\sigma_x^{w'}]$ we know that α is satisfied by $\langle \mathcal{M}'_\omega, w' \rangle$. Moreover, by Theorem 2.3 and the assumption that α is invariant for d-bisimulation, formula α is satisfied by $\langle \mathcal{M}_\omega, w \rangle$ as well. By the elementary extension, we obtain $\mathcal{M} \models \alpha[\sigma_x^{w}]$ that entails the claim (ii). Consequently, the proof is completed. ∎

Just as with SML, the key model-theoretic argument using saturation needed special care, but now with new modifications matching the above translation of $\mathsf{S_dML}$ (cf. Aucher et al., 2018).

2.4.3 Exploring expressive power

So far, we have already been able to show whether or not a first-order property belongs to the fragment identified by Theorem 2.4. In this section, we show several concrete examples, which will also present a comparison between $\mathsf{S_dML}$ and SML with respect to their expressive power.

Example 2.4: Consider the first-order property $\alpha_1(x)$ 'The current point is irreflexive and has successors, each of which only has access to the current point', i.e., $\alpha_1(x) := \neg Rxx \wedge \exists y Rxy \wedge \forall y(Rxy \to Ryx \wedge \forall z(Ryz \to z \equiv x))$. From Example 2.3, we know that this property is not invariant for d-bisimulation. For instance, formula $\alpha_1(x)$ is true at state w_1 in $\mathcal{M}_1$ but fails at v in $\mathcal{M}_2$. Thus this property is not definable in $\mathsf{S_dML}$.

Interestingly, the result may be quite different if we change the first-order property in Example 2.4 slightly, say,

Proposition 2.2: The first-order property $\alpha_1^+(x)$ 'The current point is irreflexive and has successors, each of which is a dead end or only has access to dead ends and the current point', i.e., $\alpha_1^+(x) := \neg Rxx \wedge \exists y(Rxy \wedge \neg \exists z Ryz) \wedge \exists y(Rxy \wedge \exists z Ryz) \wedge \forall y(Rxy \to$

$\neg\exists z Ryz \vee (Ryx \wedge \exists z(Ryz \wedge \neg\exists u Rzu) \wedge \forall z(Ryz \to z \equiv x \vee \neg\exists u Rzu)))$, is definable in logic $\mathsf{S_dML}$.

Proof Consider the following formulas of $\mathsf{S_dML}$:

$$(B_1) \qquad \Diamond\Box\bot \wedge \Diamond\Diamond\top$$

$$(B_2) \qquad \Box(\Diamond\top \to \Diamond\Box\bot \wedge \Diamond(\Diamond\Box\bot \wedge \Diamond\Diamond\top) \wedge \Box(\Box\bot \vee (\Diamond\Box\bot \wedge \Diamond\Diamond\top)))$$

$$(B_3) \qquad [-\Box\bot]\Box(\Diamond\Box\bot \wedge \Box(\neg\Box\bot \to \neg\Diamond\Box\bot))$$

Let $\varphi_1^+ := (B_1 \wedge B_2 \wedge B_3)$. This formula is satisfiable, say, it is true at $\langle \mathcal{M}_1, w_1 \rangle$ depicted in Figure 2.2. Let $\langle \mathcal{M}, u \rangle$ be a pointed model. It is not hard to see that $\mathcal{M}, u \vDash \varphi_1^+$ if $\mathcal{M} \vDash \alpha_1^+(x)[\sigma_x^u]$. Now assume that $\mathcal{M}, u \vDash \varphi_1^+$. Formula (B_1) states that, the current point u has some successors u_1 that are dead ends, and some successors u_2 which have successors. By (B_2), each u_2 reaches some dead end u_3, and some point u_4 which is similar to u: it has some successors which are dead ends, and some successors that also have successors. After cutting the links from node u to dead ends, from (B_3) it holds that u_2 still can see some dead ends, and that u_4 cannot reach dead ends any longer. Therefore we obtain $u_2 \neq u$ and $u_4 = u$. Consequently, $\mathcal{M} \vDash \alpha_1^+(x)[\sigma_x^u]$. So we conclude that $\mathcal{M} \vDash \alpha_1^+(x)[\sigma_x^u]$ iff $\mathcal{M}, u \vDash \varphi_1^+$ for any pointed model $\langle \mathcal{M}, u \rangle$. ∎

Through observation, we can find that the property $\alpha_1^+(x)$ expands the current point and its successors in $\alpha_1(x)$ with some successors that are dead ends. But the former one is definable in $\mathsf{S_dML}$ and the latter one is not. What is the reason for this?

Let $\langle \mathcal{M}, u \rangle$ be a pointed model that is d-bisimilar to $\langle \mathcal{M}_1, w_1 \rangle$ depicted in Figure 2.2. By Definition 2.5, we know that u can reach some dead end u_1, and some u_2 that has access to some dead ends. Except those dead ends, u_2 can also see some point u_3 that is similar to u: u_3 can reach some dead end and some node that has successors. Furthermore, after cutting the links from u to the dead ends, u_2 still can see some dead ends, but u_3 cannot reach any dead ends now. So we have $u_2 \neq u$ and $u_3 = u$. In such a way, we conclude that the property $\alpha_1^+(x)$ is invariant under d-bisimulation.

Example 2.5: Consider the property 'There exist n successors of the current point'. It is essentially not invariant for d-bisimulation. For an illustration, see Figure 2.4. Hence this property is not definable in $\mathsf{S_dML}$.

In contrast, as noted by Aucher et al. (2018), SML can count successors of the current state. Moreover, it is also expressive enough to define the length of a cycle. That is, for

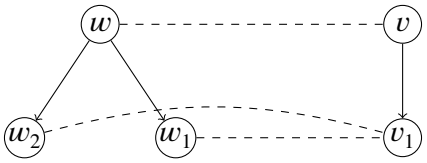


Figure 2.4 Two d-bisimilar models showing logic $\mathsf{S_dML}$ cannot count the successors of the current point. For instance, the property 'there exist 2 successors' is true at point w in the model to the left, but fails at v to the right.

each positive natural number $n \in \mathbb{N}$, there exists a SML formula φ such that, for any $\mathcal{M} = \langle W, R, V\rangle$ and $w \in W$, $\mathcal{M}, w \vDash \varphi$ iff $\langle W, R\rangle$ is a cycle of length n.[1] Is this property definable in $\mathsf{S_dML}$?

Example 2.6: Recall the two models depicted in Figure 2.3. The underlying frame of $\mathcal{M}_1$ is a cycle of length 2, while that of $\mathcal{M}_2$ is a cycle of length 1. So, logic $\mathsf{S_dML}$ cannot define the length of a cycle.

Intuitively, these differences between $\mathsf{S_dML}$ and SML stem from the features of $[-]$ and the standard sabotage modality $\blacklozenge$. In SML, each occurrence of $\blacklozenge$ in a formula deletes exactly one link. However, in $\mathsf{S_dML}$, $[-]$ operates uniformly, which blocks the logic to define the first-order properties in Example 2.5-2.6. But, the current results do not mean that $\mathsf{S_dML}$ is necessarily less expressive than SML, and the relation between these two logics remains to be clarified.

2.5 From $\mathsf{S_dML}$ to hybrid logics

While an effective first-order translation shows that validity in $\mathsf{S_dML}$ is effectively axiomatizable, it gives no concrete information about a more 'modal' complete set of proof principles. In this section, following the techniques developed by dynamic-epistemic logics (see, e.g., Baltag et al., 1998; van Benthem, 2011), we try to axiomatize $\mathsf{S_dML}$ by means of recursion axioms.

The principles for Boolean cases are as usual. However, as for $[-\varphi]\Box\psi$, there is a problem. From the typical method of recursion axioms used in dynamic-epistemic logics, we know that dynamic operators can be pushed inside standard modalities. But it fails for $\mathsf{S_dML}$, since that after pushing $[-]$ under a standard modality over successors of the current world, the model change is not local in the successors any longer and it takes place somewhere else (cf. Aucher et al., 2018).

1 For a further study of the expressivity of the sabotage-style logics (including SML), we refer to (Areces et al., 2012, 2015; Fervari, 2014) that also include comparisons of the expressivity of those logics.

Hence the principle for $[-\varphi]\Box\psi$ should illustrate the position where the change happens. To do so, a natural method is to seek help from hybrid logics (Areces and ten Cate, 2007; Blackburn and Seligman, 1995), which enable us to name nodes and specific edges in a model. This intuition is in line with the results presented in (van Benthem et al., 2020), which shows a complete axiomatization for a logic of stepwise point deletion with the help of hybrid operators.[1]

Precisely, we will extend $\mathsf{S_dML}$ with *nominals*, *at-operator* @ and *down-arrow operator* $\downarrow$, and the resulting logic is called $\mathsf{HS_dML}$. Let $\mathbf{P} = \{p, q, r, \cdots\}$ be a countable set of propositional atoms, and $\mathbf{N} = \{i, j, k, \cdots\}$ be a countable set of nominals disjoint from $\mathbf{P}$. The language of $\mathsf{HS_dML}$ is defined in the following way:

$$\varphi ::= i \mid p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Diamond\varphi \mid [-\varphi]\varphi \mid @_i\varphi \mid \downarrow_i \varphi$$

where $p \in \mathbf{P}$ and $i \in \mathbf{N}$. Models $\mathcal{M} = \langle W, R, V\rangle$ of $\mathsf{HS_dML}$ are defined as usual except that V now is a function from $\mathbf{P}\cup\mathbf{N}$ to $\mathcal{P}(W)$. In particular, for any nominal i, the valuation $V(i)$ is a singleton set. The truth condition for nominals is the following:

$$\mathcal{M}, w \vDash i \Leftrightarrow V(i) = \{w\}.$$

Truth conditions for propositional atoms, $\neg$, $\wedge$, $\Diamond$ and $[-\ ]$ are the same as those defined in Definition 2.2. Besides, given a nominal i and a formula φ, formula $@_i\varphi$ states that *φ is true at the point named by i*. Formally, it is defined by the following clause:

$$\mathcal{M}, w \vDash @_i\varphi \Leftrightarrow \mathcal{M}, v \vDash \varphi \text{ where } \{v\} = V(i).$$

Finally, formula $\downarrow_i \varphi$ says that *after naming the evaluation point i formula φ holds*, and its truth condition is defined as follows:

$$\langle W, R, V\rangle, w \vDash \downarrow_i \varphi \Leftrightarrow \langle W, R, V_i^w\rangle, w \vDash \varphi$$

where $V_i^w(i) = \{w\}$, and $V_i^w(j) = V(j)$ when $i \neq j$. In what follows, let us denote by $\mathcal{H}(\downarrow)$ the hybrid logic without operator $[-\ ]$. Now, with formulas of the form $\downarrow_i \Diamond \downarrow_j \varphi$, we can manipulate links by naming pairs of points (see Areces et al., 2016).

1 Also suggested in (Areces et al., 2018; Aucher et al., 2018), extending sabotage-style logics with hybrid operators may be an interesting method to axiomatize those logics.

2.5.1 $\mathsf{S_dML}$ and hybrid logics

As a warm-up, we briefly discuss the relation between $\mathsf{S_dML}$ and hybrid logics. In particular, the following translation illustrates that $\mathsf{S_dML}$ can be reduced to $\mathcal{H}(\downarrow)$. Similar to the standard translation defined in Definition 2.3, a finite sequence O will be used.

Definition 2.7: Let O be a finite sequence of ordered pairs consisting of nominals and $\mathcal{L}_d$-formulas, denoted with $\langle i_0, \psi_0\rangle; \cdots; \langle i_m, \psi_m\rangle; \cdots; \langle i_n, \psi_n\rangle$ $(0 \leqslant m \leqslant n)$. The *hybrid translation* $T^O : \mathcal{L}_d \to \mathcal{H}(\downarrow)$ is recursively defined as follows:

$$
\begin{aligned}
T^O(p) &= p \\
T^O(\top) &= \top \\
T^O(\neg\varphi) &= \neg T^O(\varphi) \\
T^O(\varphi_1 \wedge \varphi_2) &= T^O(\varphi_1) \wedge T^O(\varphi_2) \\
T^O(\Diamond\varphi) &= \downarrow_i \Diamond(\neg(@_i i_0 \wedge T^{\langle i,\bot\rangle}(\psi_0)) \wedge \\
&\qquad \bigwedge_{0\leqslant m\leqslant n-1} \neg(@_i i_{m+1} \wedge T^{\langle i_0,\psi_0\rangle;\cdots;\langle i_m,\psi_m\rangle}(\psi_{m+1})) \wedge T^O(\varphi)) \\
T^O([-\psi]\varphi) &= \downarrow_i T^{O;\langle i,\psi\rangle}(\varphi)
\end{aligned}
$$

where i is a nominal has not been used yet in the translation.

In the inductive clauses, formula $\Diamond\varphi$ becomes a $\mathcal{H}(\downarrow)$-formula saying that the current state, named by i, has access to a point satisfying the translation of φ, and that the link is not deleted by $[-\,]$ indexed in O. The translation of $[-\psi]\varphi$ illustrates that the translation of φ now should be taken with respect to $O; \langle i, \psi\rangle$, and that the result is true at i. Generally speaking, the truth value of a $\mathcal{H}(\downarrow)$-formula at the evaluation point may depend on the initial valuation of nominals occurring in it. However, this is not problematic: by Definition 2.7, for each $\varphi \in \mathcal{L}_d$, $T^{\langle i,\bot\rangle}(\varphi)$ yields a $\mathcal{H}(\downarrow)$-formula with at most one nominal i that is unbounded by $\downarrow$, but no points can satisfy property $\bot$ no matter what the initial valuation of i is. Now we show the correctness of Definition 2.7.

Theorem 2.5: Let φ be a formula of $\mathcal{L}_d$. For any pointed model $\langle \mathcal{M}, w\rangle$, it holds that:

$$\mathcal{M}, w \vDash \varphi \Leftrightarrow \mathcal{M}, w \vDash T^{\langle i,\bot\rangle}(\varphi).$$

Proof The proof is by induction on the structure of φ. The Boolean cases are straightforward, and we only show the non-trivial cases.

(1). When φ is $\Diamond\psi$, the following equivalences hold:

$$\begin{aligned}
\mathcal{M}, w \vDash \varphi \ \text{iff}\ & \text{there exists } v \in W \text{ s.t. } Rwv \text{ and } \mathcal{M}, v \vDash \psi \\
\text{iff}\ & \text{there exists } v \in W \text{ s.t. } Rwv \text{ and } \mathcal{M}, v \vDash T^{\langle i,\bot\rangle}(\psi) \\
\text{iff}\ & \mathcal{M}, w \vDash \Diamond T^{\langle i,\bot\rangle}(\psi) \\
\text{iff}\ & \mathcal{M}, w \vDash{\downarrow_j}\, \Diamond(\neg(@_j i \wedge T^{\langle j,\bot\rangle}(\bot)) \wedge T^{\langle i,\bot\rangle}(\psi)) \\
\text{iff}\ & \mathcal{M}, w \vDash T^{\langle i,\bot\rangle}(\varphi)
\end{aligned}$$

The first equivalence holds by the semantics of $\mathrm{SM}_d\mathrm{L}$. The second one follows from the inductive hypothesis. The third and fourth equivalences follow by the semantics of $\mathcal{H}(\downarrow)$. The last one holds by Definition 2.7.

(2). When φ is $[-\varphi_1]\varphi_2$, we have the following equivalences:

$$\begin{aligned}
\mathcal{M}, w \vDash [-\varphi_1]\varphi_2 \ \text{iff}\ & \mathcal{M}|_{\langle w,\varphi_1\rangle}, w \vDash \varphi_2 \\
\text{iff}\ & \mathcal{M}|_{\langle w,\varphi_1\rangle}, w \vDash T^{\langle i,\bot\rangle}(\varphi_2) \\
\text{iff}\ & \mathcal{M}, w \vDash{\downarrow_j}\, T^{\langle j,\varphi_1\rangle;\langle i,\bot\rangle}(\varphi_2) \\
\text{iff}\ & \mathcal{M}, w \vDash{\downarrow_j}\, T^{\langle i,\bot\rangle;\langle j,\varphi_1\rangle}(\varphi_2) \\
\text{iff}\ & \mathcal{M}, w \vDash T^{\langle i,\bot\rangle}(\varphi)
\end{aligned}$$

The first equivalence follows directly from the semantics of $\mathrm{SM}_d\mathrm{L}$. By the inductive hypothesis, for any $\langle \mathcal{M}_1, v\rangle$, it holds that $\mathcal{M}_1, v \vDash \varphi_2$ iff $\mathcal{M}_1, v \vDash T^{\langle i,\bot\rangle}(\varphi_2)$, so the second equivalence holds. Consequently, we have the third equivalence.[1] Since no point has the property $\bot$, the fourth one holds. Finally, the last equivalence holds by Definition 2.7.

Therefore, for each $\varphi \in \mathcal{L}_d$, it holds that $\mathcal{M}, w \vDash \varphi$ iff $\mathcal{M}, w \vDash T^{\langle i,\bot\rangle}(\varphi)$. ∎

In the way described, we can reduce $\mathsf{S_dML}$ to $\mathcal{H}(\downarrow)$. So, the latter one is essentially equivalent to logic $\mathsf{HS_dML}$. But, can we reduce $\mathcal{H}(\downarrow)$ to $\mathsf{S_dML}$? First note that the following property is definable in $\mathcal{H}(\downarrow)$:

Proposition 2.3: The property 'there exist n successors of the current point' is definable in $\mathcal{H}(\downarrow)$.

Proof We prove it by building the desired formula. Let n be a positive natural number. Consider the following $\mathcal{H}(\downarrow)$-formula:

1 It is worth noting that this step is not trivial. Precisely, we also need a lemma similar to the one used in the proof of Theorem 2.1.

$$\downarrow_i (\Diamond \downarrow_{i_1} (@_i \Diamond \downarrow_{i_2} (\cdots (@_i \Diamond \downarrow_{i_n} (@_i \Box (\underbrace{\bigvee_{1 \leqslant m \leqslant n} i_m \wedge \bigwedge_{1 \leqslant m < m' \leqslant n} \neg @_{i_m} i_{m'})) \cdots)))}_{n+2}$$

The formula states that the current point i has successors $i_1, \cdots, i_n$, that each node reachable from i must be some i_m, where $1 \leqslant m \leqslant n$, and that for any different m and m' such that $1 \leqslant m, m' \leqslant n$, i_m is distinct from $i_{m'}$. Thus, there exist n successors of the current point iff the stated hybrid formula holds at that point. ∎

But Example 2.5 showed that this property is not definable in $\mathsf{S_dML}$. Consequently, we have the following result:

Proposition 2.4: $\mathcal{H}(\downarrow)$ is more expressive than $\mathsf{S_dML}$ on models.

Therefore $\mathsf{S_dML}$ can be viewed as a fragment of $\mathcal{H}(\downarrow)$. Any hybrid logic at least as expressive as $\mathcal{H}(\downarrow)$ is more expressive than $\mathsf{S_dML}$. Even so, the hybrid translation described in Definition 2.7 suggests that it may be viable to analyze validity in the logic $\mathsf{S_dML}$ with expressive resources similar to those of $\mathcal{H}(\downarrow)$.

2.5.2 Digression on recursion axioms

One attractive format for axiomatizing logics of model change are recursion axioms in the style of dynamic-epistemic logics. As mentioned already, Boolean cases are available for $\mathsf{S_dML}$ as well. We begin with the principle for $[-]$:[1]

Proposition 2.5: Let φ, ψ and χ be $\mathcal{L}_d$-formulas. Then it holds that

$$[-\varphi][-\psi]\chi \leftrightarrow \downarrow_i [- \downarrow_j (\varphi \vee @_i[-\varphi]@_j\psi)]\chi \tag{2-6}$$

where i and j are new nominals.

Proof Let $\langle \mathcal{M}, w \rangle$ be a pointed model. We prove it by showing that $\mathcal{M}|_{\langle w,\varphi\rangle}|_{\langle w,\psi\rangle}$ and $\mathcal{M}|_{\langle w,\downarrow_j(\varphi \vee @_i[-\varphi]@_j\psi)\rangle}$ are identical, where $w \in V(i)$. Suppose not, then there must be some $v \in W$ such that $\langle w, v\rangle \in \mathcal{M}|_{\langle w,\varphi\rangle}|_{\langle w,\psi\rangle}$ and $\langle w, v\rangle \notin \mathcal{M}|_{\langle w,\downarrow_j(\varphi \vee @_i[-\varphi]@_j\psi)\rangle}$, or that $\langle w, v\rangle \in \mathcal{M}|_{\langle w,\downarrow_j(\varphi \vee @_i[-\varphi]@_j\psi)\rangle}$ and $\langle w, v\rangle \notin \mathcal{M}|_{\langle w,\varphi\rangle}|_{\langle w,\psi\rangle}$.

Now consider the first case. From $\langle w, v\rangle \notin \mathcal{M}|_{\langle w,\downarrow_j(\varphi \vee @_i[-\varphi]@_j\psi)\rangle}$, we know that $\mathcal{M}, v \vDash \varphi \vee @_i[-\varphi]@_j\psi$ where $v \in V(j)$. By $\langle w, v\rangle \in \mathcal{M}|_{\langle w,\varphi\rangle}|_{\langle w,\psi\rangle}$, it follows that $\mathcal{M}|_{\langle w,\varphi\rangle}, v \nvDash \psi$. Since $\mathcal{M}|_{\langle w,\varphi\rangle}|_{\langle w,\psi\rangle}$ is a submodel of $\mathcal{M}|_{\langle w,\varphi\rangle}$, we obtain $\langle w, v\rangle \in$

1 Actually, the principle for $[-]$ is not necessary to show a complete set of recursion axioms (cf. van Benthem, 2014).

$\mathcal{M}|_{\langle w,\varphi\rangle}$. Consequently, it holds that $\mathcal{M}, v \not\models \varphi$, thus, $\mathcal{M}|_{\langle w,\varphi\rangle}, v \models \psi$. So we have arrived at a contradiction.

Next we consider the second case. By $\langle w, v\rangle \in \mathcal{M}|_{\langle w,\downarrow_j(\varphi\vee@_i[-\varphi]@_j\psi)\rangle}$, it holds that $\mathcal{M}, v \models \neg\varphi \wedge @_i[-\varphi]@_j\neg\psi$ where $v \in V(j)$. Then we know $\langle w, v\rangle \in \mathcal{M}|_{\langle w,\varphi\rangle}$. Besides, by $\langle w, v\rangle \notin \mathcal{M}|_{\langle w,\varphi\rangle}|_{\langle w,\psi\rangle}$, we obtain $\mathcal{M}|_{\langle w,\varphi\rangle}, v \not\models \psi$ that entails a contradiction. ∎

Consider formula $\downarrow_i [- \downarrow_j (\varphi \vee @_i[-\varphi]@_j\psi)]\chi$. By the semantics, it is true at $\langle \mathcal{M}, w\rangle$ iff w is χ in $\mathcal{M}|_{\langle w,\downarrow_j(\varphi\vee@_i[-\varphi]@_j\psi)\rangle}$, where $V(i) = \{w\}$. Intuitively, the new model is obtained by removing all links from w to the φ-points and the points which are ψ after removing the links from w to φ-points. This is exactly what $[-\varphi][-\psi]\chi$ states.

We now move to the case for $\Box$. It seems like that the following result will work:

Proposition 2.6: For each $[-\varphi]\Box\psi \in \mathcal{L}_d$, the following equivalence holds:

$$[-\varphi]\Box\psi \leftrightarrow \downarrow_i \Box \downarrow_j (\neg\varphi \rightarrow @_i[-\varphi]@_j\psi) \tag{2-7}$$

where i and j are new nominals.

Proof Let $\langle \mathcal{M}, w\rangle$ be a pointed model. For the direction from left to right, we suppose for reductio that $\mathcal{M}, w \models [-\varphi]\Box\psi$ and $\mathcal{M}, w \not\models \downarrow_i \Box \downarrow_j (\neg\varphi \rightarrow @_i[-\varphi]@_j\psi)$. Then it holds that w ($\in V(i)$) has a successor v ($\in V(j)$) such that $\mathcal{M}, v \models \neg\varphi \wedge @_i[-\varphi]@_j\neg\psi$. From $\mathcal{M}, w \models [-\varphi]\Box\psi$, it follows that $\mathcal{M}|_{\langle w,\varphi\rangle}, w \models \Box\psi$. Since $\mathcal{M}, v \models \neg\varphi$, we obtain $\langle w, v\rangle \in \mathcal{M}|_{\langle w,\varphi\rangle}$. Thus it holds that $\mathcal{M}|_{\langle w,\varphi\rangle}, v \models \psi$. Moreover, $\mathcal{M}, v \models \neg\varphi \wedge @_i[-\varphi]@_j\neg\psi$ entails $\mathcal{M}, w \models [-\varphi]@_j\neg\psi$. Consequently, it holds that $\mathcal{M}|_{\langle w,\varphi\rangle}, v \models \neg\psi$, which entails a contradiction.

For the converse direction, assume that $\mathcal{M}, w \models \downarrow_i \Box \downarrow_j (\neg\varphi \rightarrow @_i[-\varphi]@_j\psi)$ and $\mathcal{M}, w \not\models [-\varphi]\Box\psi$. Then there exists $v \in W$ such that $\langle w, v\rangle \in R \setminus (\{w\} \times \|\varphi\|)$ and $\mathcal{M}|_{\langle w,\varphi\rangle}, v \models \neg\psi$. Consider the case that w and v are named by i and j respectively. It holds that $\mathcal{M}|_{\langle w,\varphi\rangle}, w \models @_j\neg\psi$, so $\mathcal{M}|_{\langle w,\varphi\rangle}, w \models @_i[-\varphi]@_j\neg\psi$. Furthermore, from $\langle w, v\rangle \in R\setminus(\{w\}\times\|\varphi\|)$, we know $\langle w, v\rangle \in R$ and $\mathcal{M}, v \models \neg\varphi$. Thus we have $\mathcal{M}, w \not\models \downarrow_i \Box \downarrow_j (\neg\varphi \rightarrow @_i[-\varphi]@_j\psi)$. This completes the proof. ∎

In formula (2-7), $\downarrow_i \Box \downarrow_j (\neg\varphi \rightarrow @_i[-\varphi]@_j\psi)$ states that for each successor v of the current point w, if v is not φ, then v is ψ after deleting all links from w to the φ-points. However, although formula (2-7) is valid, it is not the solution to an axiomatization of $\mathsf{S_dML}$: the formula of the form $@_i[-\varphi]@_j\psi$ blocks the recursion format, even though we have that:

Proposition 2.7: For any $p \in \mathbf{P}$, $\mathcal{L}_d$-formulas φ, ψ and χ, and nominal i, the following equivalences hold:

$$[-\varphi]@_i p \leftrightarrow @_i p \tag{2-8}$$

$$[-\varphi]@_i \neg\psi \leftrightarrow \neg[-\varphi]@_i \psi \tag{2-9}$$

$$[-\varphi]@_i(\psi \wedge \chi) \leftrightarrow [-\varphi]@_i\psi \wedge [-\varphi]@_i\chi \tag{2-10}$$

$$[-\varphi]@_i\Box\psi \leftrightarrow \downarrow_j @_i\Box \downarrow_k (\neg(\varphi \wedge @_i j) \rightarrow @_j[-\varphi]@_k\psi) \tag{2-11}$$

where j and k are new nominals.

Proof The validity of (2-8)-(2-10) is straightforward. We now consider formula (2-11). Let $\langle \mathcal{M}, w\rangle$ be a pointed model. From left to right, we suppose towards a contradiction that $\mathcal{M}, w \vDash [-\varphi]@_i\Box\psi$ and $\mathcal{M}, w \nvDash \downarrow_j @_i\Box \downarrow_k (\neg(\varphi \wedge @_i j) \rightarrow @_j[-\varphi]@_k\psi)$. Let u be a point such that $V(i) = \{u\}$. Then it holds that $\mathcal{M}, u \vDash \Diamond \downarrow_k (\neg(\varphi \wedge @_i j) \wedge @_j[-\varphi]@_k\neg\psi)$ where $w \in V(j)$. Therefore there exists some point v such that Ruv, $v \in V(k)$ and $\mathcal{M}, v \vDash \neg(\varphi \wedge @_i j) \wedge @_j[-\varphi]@_k\neg\psi$. By $\mathcal{M}, v \vDash \neg(\varphi \wedge @_i j)$, it holds that $\langle u, v\rangle \in \mathcal{M}|_{\langle w,\varphi\rangle}$. From $\mathcal{M}, v \vDash @_j[-\varphi]@_k\neg\psi$, we obtain $\mathcal{M}|_{\langle w,\varphi\rangle}, v \vDash \neg\psi$, which contradicts to $\mathcal{M}, w \vDash [-\varphi]@_i\Box\psi$.

From right to left, suppose that $\mathcal{M}, w \vDash \downarrow_j @_i\Box \downarrow_k (\neg(\varphi \wedge @_i j) \rightarrow @_j[-\varphi]@_k\psi)$ and $\mathcal{M}, w \nvDash [-\varphi]@_i\Box\psi$. Let u be a point such that $V(i) = \{u\}$. Then there exists some point v such that $\langle u, v\rangle \in \mathcal{M}|_{\langle w,\varphi\rangle}$ and $\mathcal{M}|_{\langle w,\varphi\rangle}, v \vDash \neg\psi$. From $\mathcal{M}, w \vDash \downarrow_j @_i\Box \downarrow_k (\neg(\varphi \wedge @_i j) \rightarrow @_j[-\varphi]@_k\psi)$, it holds that $\mathcal{M}, v \vDash @_j[-\varphi]@_k\psi$ where $w \in V(j)$ and $v \in V(k)$. Consequently, we have $\mathcal{M}|_{\langle w,\varphi\rangle}, v \vDash \psi$ which entails a contradiction.

The proof is completed. ∎

In the rest of this section, we are not going to present a solution to this issue. Actually we conjecture that there exists no a recursion axiom for $[-\varphi]\Box\psi$ in $\mathsf{HS_dML}$, which is contrasted with our initial intuition. However, given Corollary 2.1, there must be some sort of recursion axioms for it. Thus a question arises:

Open problem. Could there be a complete set of recursion axioms for $\mathsf{S_dML}$?

Through the above considerations, we understand why $\mathcal{H}(\downarrow)$ fails to do the job. In fact, there may be no easy solution, short of going to full first-order logic. All this suggests that, despite the axiomatizability in principle (as observed in Section 2.3), the structure of the logical validities of $\mathsf{S_dML}$ is computationally complex. This suspicion will be confirmed in the next section, where we prove the undecidability of the logic.

2.6 Undecidability of $\mathsf{S_dML}$

Up to now, we have already shown that logic $\mathsf{S_dML}$ is more expressive than the standard modal logic. Meanwhile, it is also a fragment of the hybrid logic $\mathcal{H}(\downarrow)$. It is well-known that the satisfiability problem for standard modal logic is decidable. In contrast, as noted by Blackburn and Seligman (1995), logic $\mathcal{H}(\downarrow)$ is undecidable. So, is $\mathsf{S_dML}$ decidable or not?

Actually, there are some decidable fragments of $\mathcal{H}(\downarrow)$. For instance, ten Cate and Franceschet (2005) show that after removing all formulas containing a nesting of $\Box$, $\downarrow$ and $\Box$, $\mathcal{H}(\downarrow)$ becomes decidable. But in this section, we will present a negative answer to the question above, i.e., the satisfiability problem for $\mathsf{S_dML}$ is undecidable. Interestingly, we will also identify the source of its high complexity. Before these results, we first show that $\mathsf{S_dML}$ lacks both the tree model property and the finite model property.

Theorem 2.6: The logic $\mathsf{S_dML}$ does not have the tree model property.

Proof Consider the following formulas:

$$(R_1) \qquad p \wedge \Diamond p \wedge \Diamond \neg p$$

$$(R_2) \qquad \Box(p \rightarrow \Diamond p \wedge \Diamond \neg p)$$

$$(R_3) \qquad [-\neg p]\Box\Box p$$

Let $\varphi_r := (R_1 \wedge R_2 \wedge R_3)$. We now show that for any $\mathcal{M} = \{W, R, V\}$ and $w \in W$, if $\mathcal{M}, w \vDash \varphi_r$, then the evaluation point w is reflexive. By (R_1), w has some p-successor(s) and some $\neg p$-successor(s). Formula (R_2) states that each its p-successor w_1 also has at least one p-successor w_2 and at least one $\neg p$-successor w_3. From (R_3) we know that after deleting all links from w to the $\neg p$-points, w_1 does not have $\neg p$-successors any longer. If node w_1 is not w, then φ_r cannot be true at w. That is to say, for each $v \in W$, if Rwv and $\mathcal{M}, v \vDash p$, then $v = w$, i.e., $R(w) \cap V(p) = \{w\}$. So if formula φ_r is true, the evaluation point must be reflexive (with at least one $\neg p$-successor). Besides, formula φ_r is true at v_1 in the model $\mathcal{M}_2$ depicted in Figure 2.2, so it is satisfiable. This completes the proof. ∎

In addition, $\mathsf{S_dML}$ also lacks the finite model property. To show this, inspired by the methods of Blackburn and Seligman (1995), we will construct a '*spy point*', i.e., a special point which has access in one step to any reachable point in the model.

Theorem 2.7: The logic $\mathsf{S_dML}$ does not have the finite model property.

Proof Let φ_∞ be the conjunction of the following formulas:

$$
\begin{aligned}
(F_1) &\quad s \wedge p \wedge \Box\neg s \wedge \Diamond p \wedge \Diamond\neg p \wedge \Box(\neg p \to \Box\bot)\\
(F_2) &\quad \Box(p \to \Diamond s \wedge \Diamond\neg s \wedge \Box p)\\
(F_3) &\quad \Box(p \to \Box(s \to \Box\neg s \wedge \Diamond\neg p))\\
(F_4) &\quad [-\neg p]\Box\Box(s \to \neg\Diamond\neg p)\\
(F_5) &\quad \Box(p \to \Box(\neg s \to \Diamond s \wedge \Diamond\neg s \wedge \Box p))\\
(F_6) &\quad \Box(p \to \Box(\neg s \to \Box(s \to \Box\neg s \wedge \Diamond\neg p)))\\
(F_7) &\quad [-\neg p]\Box\Box(\neg s \to \Box(s \to \neg\Diamond\neg p))\\
(Spy) &\quad \Box(p \to \Box(\neg s \to [-\neg s]\Box\Diamond(p \wedge \Box s)))\\
(Irr) &\quad \Box(p \to [-s]\Box\Diamond s)\\
(No\text{-}3cyc) &\quad \neg\Diamond(p \wedge [-s]\Diamond[-s]\Diamond\Diamond(\neg s \wedge \Box\neg s))\\
(Trans) &\quad \Box(p \to [-s]\Box\Box(\neg s \to [-\neg s]\Box\Diamond(\Box\neg s \wedge \Diamond\Box s)))
\end{aligned}
$$

First, the formula φ_∞ is satisfiable. For instance, it is true at point w in the model depicted in Figure 2.5. Now we show that for any $\mathcal{M} = \{W, R, V\}$ and $w \in W$, if $\mathcal{M}, w \vDash \varphi_\infty$, then W is infinite. For brevity, define $B = \{v \in W \mid v \in R(w) \cap V(p)\}$, i.e., B is the set of the p-successors of w. In the following proof, we assume that all previous conjuncts hold.

By (F_1), the evaluation point w is $(s \wedge p)$, and it cannot see any s-points. In particular, w cannot see itself. Also, w has some p-successor(s) (i.e., $B \neq \varnothing$) and some $\neg p$-successor(s) (i.e., $R(w) \setminus B \neq \varnothing$). In addition, each point in $R(w) \setminus B$ is a dead end.

From formula (F_2), we know that each element in B can see some $(s \wedge p)$-point(s) and $(\neg s \wedge p)$-point(s), but cannot see any $\neg p$-points. Hence each point in B has a successor distinct from itself.

According to formula (F_3), for any $w_1 \in B$, each its s-successor can see some $\neg p$-point(s), but cannot see any s-points.

By (F_4), after removing all links from w to $\neg p$-points, for each $w_1 \in B$, each of its s-successors w_2 has no $\neg p$-successors. Thus (F_4) shows that each $w_1 \in B$ can see point w, and that for each s-point $w_2 \in W$, if w_2 is a successor of w_1, then w_2 must be w.

Formulas (F_2)-(F_4) show the properties of the $(\neg s \wedge p)$-points accessible from the point w in one step. Similarly, formulas (F_5), (F_6) and (F_7) play the same roles as (F_2), (F_3) and (F_4) respectively, but focusing on showing the properties of the $(\neg s \wedge p)$-points

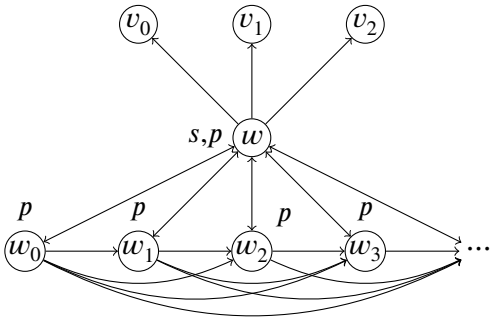


Figure 2.5 An infinite model for formula φ_∞.

that are accessible from w in two steps. In particular, (F_7) guarantees that every $(\neg s \wedge p)$-point w_1 accessible from w in two steps can also see w, and that for each s-point $w_2 \in W$, if w_2 is a successor of w_1, then w_2 must be w.

Formula (*Spy*) states that for each $(\neg s \wedge p)$-point w_1 accessible from w in two steps, after removing the links from w_1 to the $\neg s$-points, each successor w_2 of w_1 has a p-successor w_3 that only has s-successors. Furthermore, point w_2 must be s. By (F_7), we know that $w_2 = w$. In addition, by (F_2), w_3 should have some $\neg s$-successor(s) if the update induced by $[-s]$ does not take place at w_3. So we have $w_3 = w_1$. In such a way, (*Spy*) makes the evaluation point w be a spy point, and ensures that each $(\neg s \wedge p)$-point w_1 accessible from w in two steps is also accessible from w in one step. By (*Irr*), for each $w_1 \in B$, after removing $\langle w_1, w\rangle$ each successor of w_1 still can see w. Therefore, each $w_1 \in B$ is irreflexive. Besides, (*No-3cyc*) disallows cycles of length 2 or 3 in B, and (*Trans*) forces the accessibility relation R to transitively order B.

Hence B is an unbounded strict partial order, thus it is infinite and so is W. Now we have already shown that φ_∞ is satisfiable, and that for each pointed model $\langle \mathcal{M}, w\rangle$, if $\mathcal{M}, w \vDash \varphi_\infty$, then $\mathcal{M}$ is an infinite model. This completes the proof. ∎

Now, by encoding the $\mathbb{N} \times \mathbb{N}$ *tiling problem*, we show that $\mathsf{S_dML}$ is undecidable. A *tile t* is a 1×1 square, of fixed orientation, with colored edges *right(t)*, *left(t)*, *up(t)* and *down(t)*. The $\mathbb{N} \times \mathbb{N}$ tiling problem is: given a finite set T of tile types, is there a function $f : \mathbb{N} \times \mathbb{N} \to T$ such that *right(f(n,m))=left(f(n+1,m))* and *up(f(n,m))=down(f(n,m+1))*? This problem is known to be undecidable (see Harel, 1985).

Following the ideas of Blackburn and Seligman (1995), we will use three modalities $\Diamond_s$, $\Diamond_u$ and $\Diamond_r$. Correspondingly, a model $\mathcal{M} = \{W, R_s, R_u, R_r, V\}$ now has three accessibility relations. We will construct a spy point over the relation R_s. The relations R_u and R_r represent moving up and to the right, respectively, from one tile to the other.

In addition, the operator $[-\ ]$ will work in the usual way, i.e., all of the three kinds of relations should be cut if the current point has some particular successors via them.[1] Let us see the details.

Theorem 2.8: The satisfiability problem for $\mathsf{S_dML}$ with three binary accessibility relations R_s, R_u and R_r is undecidable.

Proof Let $T = \{T_1, \cdots, T_n\}$ be a finite set of tile types. For each $T_i \in T$, we use $u(T_i)$, $d(T_i)$, $l(T_i)$, $r(T_i)$ to represent the colors of its up, down, left and right edges respectively. Also, we code each tile type T_i with a fixed propositional atom t_i. Now we will define a formula φ_T such that φ_T is satisfiable iff T tiles $\mathbb{N} \times \mathbb{N}$. Consider the following formulas:

$$
\begin{array}{rll}
(M_1) & s \wedge p \wedge \Box_s \neg s \wedge \Diamond_s p \wedge \Diamond_s \neg p \wedge \Box_s(\neg p \to \Box_s \bot) & \\
(M_2) & \Box_s(p \to \Diamond_s \top \wedge \Box_s(s \wedge \Diamond_s \neg p)) & \\
(M_3) & [-\neg p]\Box_s \Box_s(s \wedge \neg \Diamond_s \neg p) & \\
(M_4) & \Box_s(p \to \Diamond_\dagger \top \wedge \Box_\dagger(\neg s \wedge p \wedge \Diamond_s \top \wedge \Box_s(s \wedge \Diamond_s \neg p))) & \dagger \in \{u, r\} \\
(M_5) & [-\neg p]\Box_s \Box_\dagger \Box_s \neg \Diamond_s \neg p & \dagger \in \{u, r\} \\
(M_6) & \Box_s(p \to \Box_\dagger(\Diamond_u \top \wedge \Diamond_r \top \wedge \Box_u(\neg s \wedge p) \wedge \Box_r(\neg s \wedge p))) & \dagger \in \{u, r\} \\
(M_7) & \Box_s(p \to [-s]\Box_\dagger(\Diamond_s s \wedge \neg \Diamond_\dagger \neg \Diamond_s s)) & \dagger \in \{u, r\} \\
(Spy) & \Box_s(p \to \Box_\dagger[-\neg s]\Box_s \Diamond_s(p \wedge \Box_u \bot \wedge \Box_r \bot)) & \dagger \in \{u, r\} \\
(Func) & \Box_s(p \to [-s]\Box_\dagger[-\neg s]\Diamond_s \Diamond_s(p \wedge \neg \Diamond_s s \wedge \Diamond_\dagger \top \wedge & \\
 & \Box_\dagger(\Box_u \bot \wedge \Box_r \bot)) & \dagger \in \{u, r\} \\
(No\text{-}UR) & \Box_s(p \to [-s]\Box_u \Box_r \Diamond_s s \wedge [-s]\Box_r \Box_u \Diamond_s s) & \\
(No\text{-}URU) & \Box_s(p \to [-s]\Box_u \Box_r \Box_u \Diamond_s s) & \\
(Conv) & \Box_s(p \to [-s]\Diamond_u [-s]\Diamond_r [-\neg s]\Diamond_s \Diamond_s(p \wedge \neg \Diamond_s s \wedge & \\
 & \Box_r(\Diamond_u \top \wedge \Diamond_r \top) \wedge \Diamond_u \neg \Diamond_s s \wedge \Diamond_r \Diamond_u(\Box_u \bot \wedge \Box_r \bot))) & \\
(Unique) & \Box_s(p \to \bigvee\limits_{1 \leqslant i \leqslant n} t_i \wedge \bigwedge\limits_{1 \leqslant i < j \leqslant n} (t_i \to \neg t_j)) & \\
(Vert) & \Box_s(p \to \bigwedge\limits_{1 \leqslant i \leqslant n} (t_i \to \Diamond_u \bigvee\limits_{1 \leqslant j \leqslant n,\ u(T_i) = d(T_j)} t_j)) & \\
(Horiz) & \Box_s(p \to \bigwedge\limits_{1 \leqslant i \leqslant n} (t_i \to \Diamond_r \bigvee\limits_{1 \leqslant j \leqslant n,\ r(T_i) = l(T_j)} t_j)) &
\end{array}
$$

1 There is also no problem if we use three kinds of dynamic operators that correspond to the three accessibility relations respectively. In the proof of Theorem 2.8, these three kinds of links are disjoint.

Define φ_T as the conjunction of the formulas above. Let $\mathcal{M} = \{W, R_s, R_u, R_r, V\}$ be a model and $w \in W$ such that $\mathcal{M}, w \vDash \varphi_T$. We show that $\mathcal{M}$ is a tiling of $\mathbb{N} \times \mathbb{N}$. For brevity, define $G = \{v \in W \mid v \in R_s(w) \cap V(p)\}$ where $R_s(w) = \{v \in W \mid R_s wv\}$, and we will use its elements to represent the tiles. In the following proof, we also assume that all previous conjuncts hold.

Formula (M_1) is similar to (F_1) occurring in the proof of Theorem 2.7, except that (M_1) only concerns the relation R_s.

By (M_2), each tile w_1 has some successor(s) via the relation R_s, and each such successor w_2 is $(s \wedge p)$ and also has some $(\neg s \wedge \neg p)$-successor(s) via R_s. It is worthy to note that formulas (M_1) and (M_2) illustrate that R_s is irreflexive.

Formula (M_3) ensures that each tile w_1 can see w via R_s, and that for each $(s \wedge p)$-point $w_2 \in W$, if w_2 is accessible from w_1 via R_s, then $w_2 = w$.

(M_4) states that each tile has some successor(s) via R_u and some successor(s) via R_r. Furthermore, each point accessible from a tile via R_u or R_r is very similar to a tile: it is $(\neg s \wedge p)$, and has some $(s \wedge p)$-successor(s) w_1 via relation R_s where each w_1 can see some $(\neg s \wedge \neg p)$-point(s) via R_s.

By formula (M_5), each $w_1 \in W$ accessible from a tile via R_u or R_r can see w by R_s. Also, for each $(s \wedge p)$-point $w_2 \in W$, if it is accessible from w_1 via R_s, then $w_2 = w$.

Formula (M_6) ensures that each $w_1 \in W$ accessible from some tile via R_u or R_r also has some successor(s) via R_u and some successor(s) via R_r. Besides, each its successor via R_u or R_r is $(\neg s \wedge p)$.

From formula (M_7), it follows that both R_u and R_r are irreflexive and asymmetric.

By (*Spy*), we know that the evaluation point w is a spy point via the relation R_s.

Note that formula (M_4) shows that each tile has some tile(s) above it and some tile(s) to its right. Now, with (*Func*), we have that each tile has exactly one tile above it and exactly one tile to its right.

By (*No-UR*), no tile can be above/below as well as to the left/right of another tile. Formula (*No-URU*) disallows cycles following successive steps of the R_u, R_r, and R_u relations, in this order. Furthermore, (*Conv*) ensures that the tiles are arranged as a grid.

Formula (*Unique*) guarantees that each tile has a unique type. (*Vert*) and (*Horiz*) force the colors of the tiles to match properly.

Thus we conclude that $\mathcal{M}$ is indeed a tiling of $\mathbb{N} \times \mathbb{N}$.

Next we show the other direction required for our proof. Suppose the function f :

$\mathbb{N} \times \mathbb{N} \to T$ is a tiling of $\mathbb{N} \times \mathbb{N}$. Define a model $\mathcal{M} = \{W, R_s, R_u, R_r, V\}$ as follows:

$$
\begin{aligned}
W &= (\mathbb{N} \times \mathbb{N}) \cup \{w, v\} \\
R_s &= \{\langle w, v\rangle\} \cup \{\langle w, x\rangle \mid x \in \mathbb{N} \times \mathbb{N}\} \cup \{\langle x, w\rangle \mid x \in \mathbb{N} \times \mathbb{N}\} \\
R_u &= \{\langle\langle n, m\rangle, \langle n, m+1\rangle\rangle \mid n, m \in \mathbb{N}\} \\
R_r &= \{\langle\langle n, m\rangle, \langle n+1, m\rangle\rangle \mid n, m \in \mathbb{N}\} \\
V(s) &= \{w\} \\
V(p) &= \{w\} \cup (\mathbb{N} \times \mathbb{N}) \\
V(t_i) &= \{\langle n, m\rangle \in \mathbb{N} \times \mathbb{N} \mid f(\langle n, m\rangle) = T_i\}, \text{ for each } i \in \{1, \cdots, n\} \\
V(q) &= \varnothing, \text{ for any other propositional atoms } q
\end{aligned}
$$

In particular, w is a spy point in $\mathcal{M}$. By construction, we know that $\mathcal{M}, w \vDash \varphi_T$. ∎

It is important to notice that the three relations used in the proof above can be reduced to one, by using an argument analogous to the one presented in (Hoffmann, 2015) which uses propositional symbols to appropriately encode the relations R_u and R_r.[1] Thus, perhaps surprisingly, given the simple-looking syntax and semantics of $\mathsf{S_dML}$, the complexity of its logic is high. What is the reason for this high complexity, as contrasted with decidability of dynamic-epistemic logics of link deletion (cf. van Benthem and Liu, 2007)? For SML, the reason offered by Aucher et al. (2018) is the stepwise nature of link deletion, and this is confirmed by the result of van Benthem et al. (2020) showing how a very simple stepwise variant of public announcement logic is undecidable. However, our case is different, since links are cut in a uniform definable way: the only remaining potential culprit is then the locality.

To see the effects of this feature, recall the above formula (2-7). We already saw in Section 2.5.2 that a formula of the form $@_x[-\varphi]@_y\psi$ blocks the recursion format. In contrast, let us consider a global version $\mathsf{S_d^gML}$ of $\mathsf{S_dML}$. Now formula $[-\varphi]\psi$ states that ψ is true at the evaluation point after deleting all links to φ-worlds, i.e.,

$$\langle W, R, V\rangle, w \vDash [-\varphi]\psi \Leftrightarrow \langle W, R \setminus \{\langle s, t\rangle \in R \mid \mathcal{M}, t \vDash \varphi\}, V\rangle, w \vDash \psi.$$

1 More directly, we can also show the undecidability of $\mathsf{S_dML}$ (with one accessibility relation) by reduction from other undecidable logics. Closely related to our work, Areces et al. (2018) prove the undecidability results for several sabotage-style logics, such as SML and its local variant, by reductions from an undecidable memory logic.

Given the global change made in this semantics, here is a valid recursion axiom for $\Box$:

$$[-\varphi]\Box\psi \leftrightarrow \Box(\neg\varphi \rightarrow [-\varphi]\psi).$$

Indeed, following the general method for modal logics of definable model change provided by van Benthem and Liu (2007), one can find a complete set of recursion axioms for $\mathsf{S}^{\mathsf{g}}_{\mathsf{d}}\mathsf{ML}$:

Proposition 2.8: The logic $\mathsf{S}^{\mathsf{g}}_{\mathsf{d}}\mathsf{ML}$ is axiomatizable and decidable.

The complexity effect of the local behavior of $\mathsf{S}_{\mathsf{d}}\mathsf{ML}$ also show at a crucial step in our proof of undecidability. In the proof of Theorem 2.8, formula (*Conv*) forces the tiles to satisfy a first-order convergence property, i.e.,

$$\forall t\forall t_1\forall t_2(R_u t t_1 \wedge R_r t_1 t_2 \rightarrow \exists t_3(R_r t t_3 \wedge R_u t_3 t_2)).$$

As noted by van Benthem (2010), this property can give logics high complexity.

By contrast, convergence is not definable in $\mathsf{S}^{\mathsf{g}}_{\mathsf{d}}\mathsf{ML}$, even though we expand the model with some extra tools (e.g., a spy point). Roughly speaking, given two tiles t_1 and t_2 that have the same properties, we still can distinguish between them with $\mathsf{S}_{\mathsf{d}}\mathsf{ML}$, say, their properties will be different after cutting some links starting from t_1; however, we cannot do this with $\mathsf{S}^{\mathsf{g}}_{\mathsf{d}}\mathsf{ML}$, since links are cut in a global way.[1]

The more general issue arising here goes beyond our specific logics of sabotage:

Open problem. Does making update operations local (world-relative) generate undecidability in general for decidable dynamic-epistemic logics?

This would provide an alternative diagnosis to the comparison of sabotage and update offered by Aucher et al. (2018), closer to the modified dynamic-epistemic logics studied by Belardinelli et al. (2017).

2.7 Related work

Graph games and logics. The work of this chapter is primarily inspired by existing work on sabotage games, sabotage modal logic and their variants (see, e.g., Aucher et al., 2015, 2018; Gierasimczuk et al., 2009; Rohde, 2005; van Benthem, 2014). So far, several properties of SML have been studied. As illustrated, the first-order translation for SML is described independently by Areces et al. (2015); Aucher et al. (2015), which together

1 From a technical point of view, to show that $\mathsf{S}^{\mathsf{g}}_{\mathsf{d}}\mathsf{ML}$ cannot define the convergence property, we need its notion of bisimulation, which is easily defined.

with Areces et al. (2012) also propose a suitable notion of bisimulation for SML that solves an important open problem mentioned by Löding and Rohde (2003a). Besides, Löding and Rohde (2003a) show that the multi-modal version of SML has a PSPACE-complete model checking problem and an undecidable satisfiability problem. These two results are improved by (Areces et al., 2012, 2015), which show that they also hold for SML. For the latest developments in sabotage modal logics, we refer the reader to (Aucher et al., 2018) that also has extensive references to current work on related modal logics for definable graph change.

Meanwhile, a number of authors have studied other graph games using matching modal logics. For instance, in poison games, originating in graph theory, instead of deleting links, a player can poison a node to make it inaccessible to her opponent. Poison games have been recently studied in the modal logics of Blando et al. (2020), using the close similarities between these systems and variants of so-called *memory logics* (Areces, 2007; Areces et al., 2011) in the hybrid tradition. In another tradition, that of Boolean network games, Thompson (2020) has proposed a logic of local fact change which can characterize Nash equilibria, providing a new way of looking at the interaction between graph games, network games and logics of control.

It remains to note that this chapter fits with the general program recently proposed by van Benthem and Liu (2020) for a much broader study of analysis and design for graph games in tandem with matching modal logics. In particular, it proposes various meaningful new games, and identifies general questions behind the match between logics and games.

Logics with model modifiers. In addition to SML, our logic $\mathsf{S_dML}$ is also closely related to other logics with model modifiers. Recently, an important series of research is the work on *relation-changing logics* (e.g., Areces et al., 2012, 2015, 2016, 2018; C. Areces and R. Fervari and G. Hoffmann, 2014), which include modalities to swap, delete or add links. It is worth noting that they also investigate a special type of local SML, whose dynamic operator refers to a model transition that cuts a link from the current state and then evaluates a formula at the target of the deleted arrow. A general view on these logics is presented in (Fervari, 2014), which studies various meta-properties, such as expressive power, complexity, tableaux methods and their relations with dynamic-epistemic logics.

Different from relation-changing logics, van Benthem et al. (2020) develop a logic of stepwise point deletion. This work is helpful to understand the complexity jumps be-

tween dynamic-epistemic logics of model transformations and logics of freely chosen graph changes recorded in current memory. Moreover, techniques developed by van Benthem et al. (2020) also shed light on the long-standing open problem of how to axiomatize sabotage-style modal logics and related ones.

Finally, more akin to the above-mentioned (Thompson, 2020), Aucher et al. (2009) study global and local modifiers that update the valuation at the evaluation point, and show that adding all those local modifiers dramatically increases the expressive power of the logic without them.

Dynamic-epistemic logics. Throughout the chapter, dynamic-epistemic logics (see, e.g., Baltag et al., 1998; van Benthem, 2011) have been used as a decidable contrast to our systems. Technically, $\mathsf{S_dML}$ has resemblances to several recent logics for local announcements. Belardinelli et al. (2017) introduce a logic to characterize both global and local announcements. Similar to our set-up, it has definable updates of links, but there is a crucial difference: the logic is more expressive than public announcement logic, but its satisfiability problem still is decidable. It is important to recognize that the decidability result cannot be treated as a negative answer to the open problem proposed in Section 2.6, although it is noted by Belardinelli et al. (2017) that locality is also a distinguishing feature of that framework. Roughly speaking, the accessibility relations studied in that paper are equivalent relations and locality is defined with respect to agents, which is quite different from our work.

Hybrid logics. Another highly relevant line of research for this chapter is hybrid logics (Areces and ten Cate, 2007; Blackburn and Seligman, 1995), an area from which we have taken several basic techniques. As far as we know, Blackburn and Seligman (1995) are the first to present the method of constructing a spy point, the main tool that is used to prove the undecidability of our logic $\mathsf{S_dML}$. Besides, Areces et al. (2016) show how relation-changing logics such as SML can be seen as fragments of hybrid logics, and identify various decidable fragments of those logics with the help of hybrid translations. This fits with our findings in Section 2.5.1. Conversely, as mentioned in Section 2.6, Areces et al. (2018) prove the undecidability results for a number of relation-changing logics by reduction from memory logic. However, no such translations from hybrid logic or memory logic into $\mathsf{S_dML}$ exist in an obvious way. Finally, Hansen (2011) merges a hybrid logic with public announcement logic. Different from the operator $[-\ ]$ in $\mathsf{S_dML}$, the

announcement modality there operates in a global way, making it possible to axiomatize the logic by means of recursion axioms.

Reactive modal logic. Finally, it is important to recognize that $\mathsf{S_dML}$ shares a typical feature with the *reactive modal logic* (Gabbay, 2008, 2013), i.e., the accessibility relation of models is changed during the interpretation process of formulas. However, there are also various typical differences. For instance, unlike our case, the update studied by Gabbay (2008, 2013) is not definable. In addition, the language of reactive modal logic is the standard modal language $\mathcal{L}_{\Box}$, but formulas are evaluated in the so called *reactive Kripke models* instead of standard relational models. In contrast, the language $\mathcal{L}_d$ of $\mathsf{S_dML}$ is an extension of $\mathcal{L}_{\Box}$, and the truth conditions for Boolean connectives and $\Box$ are as usual. More importantly, the level of updates in our work is different from that of Gabbay (2008, 2013). In our proposal, language $\mathcal{L}_d$ is equipped with a dynamic operator $[-\,]$ to encode the desired changes. Precisely, for any pointed model $\langle \mathcal{M}, w\rangle$ and formula $[-\varphi]\psi$, the update $[-\varphi]$ produces another model $\mathcal{M}|_{\langle w,\varphi\rangle}$ and formula ψ now is evaluated at w in the new model. Different from this, reactive Kripke models extend standard relational models with a special kind of links from points or links to links, which get activated during the modal process and change the model. Therefore, deletion in our work is a metalevel notion, while it is brought into the object level in (Gabbay, 2008, 2013).[1]

2.8 Summary and future work

Summary. In the chapter, we started with some basic observations on interactions of agents in our social reality, from a particular perspective: the multi-agent communication with agents deeply at odds, in that they try to change the background in which their interactions are performed. To characterize those situations formally, we appealed to the techniques of graph games and explored a notion of definable sabotage games, where players could remove links with an explicit definition. As illustrated, different readings of the destruct actions would enable the framework to capture many interesting interactions in various social contexts.

Moreover, our games had a natural relation with logics. As we have shown, our logical system $\mathsf{S_dML}$ contained proper operators matching with players' actions, and furthermore, it was also helpful to define their winning positions. The logic, as a formal tool

1 Similar to our case, all kinds of updates studied in the work on relation-changing logics (e.g., SML) are also metalevel notions.

to reason the games, would throw light on many crucial problems involved with the social situations mentioned above. Say, when some agents in a social network break off their relations with others step by step, does an agent of particular importance remain reachable via intermediate ones? After indicating the applications of the logic, we also explored many important properties of the logic, including:

- We presented a first-order translation for the logic, showed a characterization theorem with regard to a novel notion of definable sabotage bisimulation.
- We probed options for a perspicuous axiomatization for $\mathsf{S_dML}$ using recursion axioms in an extended hybrid language.
- We proved that the logic does not have the finite model property and its satisfiability problem is undecidable. Moreover, we identified the feature causing the high complexity.

Together, these results show that our new perspective on social interaction as modeled by graph games has some mathematical substance, and allows us to investigate the potential sources of complexity in reasoning about social scenarios. Moreover, we believe that this tool for analysis might combine well with those offered by other formal models of agency, in particular, those offered by game theory.

Further research. Immediate technical open problems for our logic $\mathsf{S_dML}$ resemble those in the literature for SML. For instance, we would like to have a good Hilbert-style proof theory, which may perhaps be found by analyzing semantic tableaux for $\mathsf{S_dML}$. Another open problem is the axiomatization and complexity of the schematic validities of our language, that remain valid under arbitrary substitutions for atoms.

In terms of generality, one would like to establish the precise connections between our logic $\mathsf{S_dML}$ and other modal logics for graph games in the cited literature. For instance, the difference in expressive power that we noted in Section 2.4 between $\mathsf{S_dML}$ and SML does not preclude the existence of faithful embedding either way.[1]

As a final technical issue, we mentioned the contrast between locality and stepwise link deletion as sources of undecidability, discussed in Section 2.6. One could also merge these in a stepwise version of our logic, denoted $\mathsf{S_d^sML}$. Clearly, its validities are different from those of $\mathsf{S_dML}$: for instance, $[-]$ is no longer self-dual. Our methods from Section 6 should also be able to prove its undecidability, but we have not yet been able to do so.

1 Given the various new translations between logics of point deletion and link deletion in (van Benthem et al., 2021a), there may also be more to the connections between our logics and the earlier system MLSR) of stepwise point removal in van Benthem et al. (2020).

We end by stepping back to reality. In our introduction, we mentioned social networks, where adding links (gaining friends or neighbors) is as important as deleting links (losing friends or neighbors). A connection between our logic and existing logics for social networks (Liu et al., 2014), now adding various graph games played over these, would be a natural next step for the modeling techniques offered in this chapter.

Indeed, this chapter can be considered a case study for the general program put forward in van Benthem and Liu (2020) for a much broader study of analysis and design for graph games in tandem with matching modal logics. In particular, this program extends to a much broader range of games, including families of parlor games in actual use, and it includes the explicit design of various meaningful new games, while identifying general questions behind the match between game design and logic design. Of the steps towards greater realism that arise in such a program, we mention the possibility of *more complex independent goals* for players in social scenarios than we have considered, which may line up to some extent, though perhaps not completely. An equally urgent extension is the introduction of *imperfect information* and the crucial role of knowledge and ignorance in scenarios where players cannot perfectly observe each other's moves. In general, such games may have only probabilistic equilibria, and our logics would have to acquire interfaces with probability.

Chapter 3 Interactions in learning and teaching - A graph game approach

3.1 Introduction: correct learning games

The preceding chapter showed how graph games with link deletions in sabotage style provide us with a precise tool to model a number of social interactions under adverse circumstances. But the technique of link modifications presented there itself is absolutely neutral, and its very abstraction also enables us to analyze more 'positive' scenarios. This is exactly what we will show in the chapter: as we shall see, graph games with link modifications can also highlight the interactive nature of teaching and learning, and characterize a number of interesting features highlighted Example 1.2 in Chapter 1. In a scenario of learning and teaching, there are many natural questions that need to be asked. For instance, and perhaps most prominently,

- When can we say that the learner in question has succeeded in learning?
- Can the learner achieve the goal under the guidance of her teacher?

Different mathematical frameworks for such scenarios may give different answers.

As observed and elaborated in Gierasimczuk et al. (2009), the original sabotage game SG defined in earlier chapters can be used as a concise model for capturing teaching scenarios where teacher drag perhaps unwilling students to a state of knowledge on some topic at hand. In this chapter, we restrict ourselves to cases where Learner (i.e., Traveler) is *eager* to learn, and Teacher (i.e., Blocker) is *helpful*. That is, reaching the correct node is the goal of both Learner and Teacher, which roughly depicts a guided learning situation. To demonstrate the intuition behind the correspondence between SG and those scenarios, consider the following reading that is closely related to Example 1.2:

A teaching-interpretation of SG: theorem proving. In this context, the starting node intuitively is given by axioms, the goal node stands for the theorem to be proved, other nodes represent lemmas conjectured by Learner, and edges capture Learner's possible inferences between them. Inferring is represented by moving along edges. The information provided by Teacher can be treated as his feedback, i.e., removing edges to eliminate wrong inferences. The success condition is given by the winning condition: the learning

process has been successful if Learner reaches the goal node, i.e., proving the theorem.[1]

Following the direction, we would like to enrich the notion of sabotage games with further ingredients that are able to provide insights into the questions involving Example 1.2 in Chapter 1 and those above, and meanwhile capture more realistic features of learning processes. In particular, we are going to highlight the following aspects of the interactions of Learner and Teacher:

- There are correct inferences as well as incorrect ones, but Learner cannot distinguish the difference between them.
- Teacher does not have to act in each round: in fact, the requirement that Teacher needs to act in each round may cause undesired outcomes, which can be easily illustrated by a simple example of sabotage games. Also, it is not necessary that all correct inferences between different hypotheses have already been conjectured by Learner: during the process of learning, 'possibilities may also be ignored due to the more questionable practice if assuming that one of the theories under consideration must be true. And complexity can come to be ignored through convention or habit' (see pp. 260 Kelly et al., 1997). In this case, Teacher might point out the facts ignored.
- Links removed represent mistakes. So, whether or not a link deleted has occurred in Learner's current proof (i.e., the current process of the learning) matters. If the proof includes a wrong inference, any further steps of the proof should not make sense. However, if a potential transition having not occurred in the proof is wrong, Learner can continue with her current position.
- The well-known Gettier cases (Gettier, 1963) indicate that reaching the right conclusion may also be unreliable: Learner should reach the goal in a correct way. So, a solid success condition for learning should ask Learner to come to the conclusion in a coherent way.

To capture these features, we are going to propose a new framework, called *correct learning games* CLG. The notion differs from SG on several accounts. Before its definition, let us define some auxiliary notions.

Let $S = \langle w_0, w_1, \cdots, w_n \rangle$ be a non-empty, finite sequence. We denote by $e(S)$ the last element of S, and $S; v$ the sequence extending S with the element v. Define $Set(S) :=$

1 The interpretation can be easily adapted to characterize other situations of learning. For the general correspondence between SG and learning models, we refer to (Gierasimczuk et al., 2009).

$\{\langle w_0, w_1\rangle, \langle w_1, w_2\rangle, \cdots, \langle w_{n-1}, w_n\rangle\}$, denoting the set of links occurring in the sequence. When S is a singleton, $Set(S) := \varnothing$. Also, for any $\langle w_i, w_{i+1}\rangle \in Set(S)$, $S|_{\langle w_i, w_{i+1}\rangle} := \langle w_0, w_1, \cdots, w_u\rangle$, where $\langle w_u, w_{u+1}\rangle = \langle w_i, w_{i+1}\rangle$ and $\langle w_u, w_{u+1}\rangle \neq \langle w_j, w_{j+1}\rangle$ for any $j < i$. Intuitively, $S|_{\langle w_i, w_{i+1}\rangle}$ is obtained by deleting all elements occurring after w_u from S, where $\langle w_u, w_{u+1}\rangle$ is the first occurrence of $\langle w_i, w_{i+1}\rangle$ in S. Say, when $S = \langle a, b, c, a, b\rangle$, we have $S|_{\langle a,b\rangle} = \langle a\rangle$. Now let us introduce CLG.

Definition 3.1: A *correct learning game* CLG $\langle W, R_L, R_T, \langle s\rangle, g\rangle$ is given by a graph $\langle W, R_L, R_T\rangle$, the starting node s and the goal node g. A position of the game is a tuple $\langle R_L^i, S^i\rangle$. The initial position $\langle R_L^0, S^0\rangle$ is given by $\langle R_L, \langle s\rangle\rangle$. Round $n+1$ from position $\langle R_L^n, S^n\rangle$ is as follows: first, Learner moves from $e(S^n)$ to any of its R_L-successors s'; then Teacher does nothing or acts out one of the following three choices:

(a). Extend R_L^n with some $\langle v, v'\rangle \in R_T$;

(b). Transfer $S^n; s'$ to $(S^n; s')|_{\langle v,v'\rangle}$ by cutting $\langle v, v'\rangle$ from $Set(S^n; s') \setminus R_T$;

(c). Delete some $\langle v, v'\rangle \in (R_L^n \setminus R_T) \setminus Set(S^n; s')$ from R_L^n.

The resulting position, denoted $\langle R_L^{n+1}, S^{n+1}\rangle$, is $\langle R_L^n, S^n\rangle$ (when Teacher does nothing), $\langle R_L^n \cup \{\langle v, v'\rangle\}, S^n; s'\rangle$ (when he chooses (a)), $\langle R_L^n \setminus \{\langle v, v'\rangle\}, (S^n; s')|_{\langle v,v'\rangle}\rangle$ (if he acts as in (b)), or $\langle R_L^n \setminus \{\langle v, v'\rangle\}, S^n; s'\rangle$ (if he chooses (c)). It ends if Learner arrives at g through an R_T-path $\langle s, \cdots, g\rangle$ or cannot make a move, with both players winning in the former case and losing in the latter.

Intuitively, the clause for Learner illustrates that she cannot distinguish the links starting from the current position. The sequence S^i is her current learning process, which may include mistakes; R_L represents Learner's possible inferences; and R_T is the correct inferences. For any position $\langle R_L^n, S^n\rangle$, $Set(S^n) \subseteq R_L^n$. Besides, (b) and (c) focus on the case where Teacher eliminates wrong transitions, but there is an important difference. Action (b) concerns the case where Teacher gives Learner a counterexample to show that she has gone wrong somewhere in her current process, so Learner should move back to the conjecture right before the wrong transition. In contrast, (c) illustrates that Teacher eliminates a wrong transition conjectured having not occurred in Learner's process yet, therefore it does not modify Learner's current process.

From the winning condition, we know that both the players cooperate with each other. It is important to recognize that Learner's action does not conflict with her cooperative nature: to achieve the goal, she tries to move in each round. For an example of CLG, see

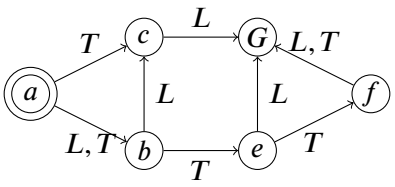


Figure 3.1 A CLG. In this graph, R_L is labelled with 'L' and R_T with 'T'. The starting node is a and the goal node is G. We show that players have a winning strategy by depicting the game to play out as follows. Learner begins by moving along the only available edge to node b. Teacher in his turn can make $\langle e, f\rangle$ 'visible' to Learner by adding it to R_L. Then, Learner proceeds to move along $\langle b, c\rangle$, and Teacher extends $\langle b, e\rangle$ to R_L. Afterwards, Learner continues on the only option $\langle c, G\rangle$. Although she now has already arrived at the goal node, her path $\langle a, b, c, G\rangle$ is not an R_T-sequence. So, Teacher can remove $\langle b, c\rangle$ moving Learner back to node b. Next, Learner has to move to e, and Teacher can delete $\langle e, G\rangle$ from R_L. Finally, Learner can arrive at the goal node G in 2 steps with Teacher doing nothing. Now we have $Set(\langle a, b, e, f, G\rangle) \subseteq R_T$, so they win.

Table 3.1 Correspondence between theorem proving and correct learning games.

Theorem proving	**Correct learning games**
Axioms	Starting node
Theorem	Goal node
Lemmas conjectured by Learner	Other states except the starting state and the goal state
Learner's possible inference from a to b	R_L-edge from a to b
Correct inference from a to b	R_T-edge from a to b
Inferring b from a	Transition from a to b
Proof for a	R_L-sequence from the starting node to a
Correct proof for a	R_L-sequence S from the starting node to a and $Set(S) \subseteq R_T$
Giving a counterexample to the inference from a to b in the proof S	Modifying S to $S|_{\langle a,b\rangle}$ ($\langle a, b\rangle \in Set(S)$)
Giving a counterexample to the conjectured inference from a to b not in the proof S	Deleting $\langle a, b\rangle$ from R_L ($\langle a, b\rangle \notin Set(S)$)
Pointing out a potential inference from a to b not conjectured by Learner before	Extending R_L with $\langle a, b\rangle$

Figure 3.1. The correlation between the situation of theorem proving and CLG is shown in Table 3.1.

Remark 3.1: The interpretation of CLG in Table 3.1 can be easily adapted to characterize other paradigms in formal learning theory, such as language learning and scientific inquiry. More generally, any single-agent games, such as solitaire and computer games,

can be converted into CLG. Say, the player (Learner) does not know the correct moves well, but she knows the starting position and the goal position, and has some conjectures about the moves of the game. Besides, she can be taught by Teacher: she just attempts to play it, while Teacher instructs her positively (by revealing more correct moves) or negatively (by pointing out incorrect moves, in which case Learner may have to be moved back to the moment previous to the first incorrect move, if she made any).

Finally, we end this part by a preliminary comparison of CLG and SG.

First, note that Learner in a SG can win only if the graph contains a sequence of edges from the starting node to the goal. Similarly, in a CLG, players cannot win when there exists no R_T-path from the starting node to the goal node. From the perspective of learning, both these two conditions are reasonable: the interaction between Learner and Teacher makes sense only when the goal is learnable. However, it is important to recognize that in both SG and CLG, the existence of such a path cannot guarantee their winning.

Also, there are several notable differences between SG and CLG. In a SG, Learner knows the underlying graph well, and is always on one of the paths with which she can finally arrive at the goal (if they exist). Therefore, she has the ability to move to a suitable node in the next round. In contrast, the player in a CLG does not have this ability: all R_L-links starting from her current position looks 'the same' from her perspective, and she is not able to guarantee that her movements are always the good ones (even though sometimes she may move to some 'good' nodes by chance). As to Teacher, compared with that of SG, the player in CLG is more powerful: he now is not only able to remove links, but also able to add new edges to the graph. However, from another aspect, the ability of Teacher in CLG is more restrictive as well: he can only delete the wrong translations from the graph.

An interesting issue worth studying is the precise relationship between SG and CLG. One observation involving this is as follows. Given a SG including a path with which the players can win, we can build a CLG by labelling the links of the path with both 'L' and 'T' and all others in the initial graph with 'T' only. In the CLG constructed, with the same path as that of the initial SG, the players can also win: Learner just moves (in each round there exists only one R_L-successor of her current position), and Teacher does not need to do anything. The observation is restrictive, but it seems there does not exist an obvious way to encode SG into CLG generally. We leave this for future work.

In the remainder of this chapter, we will study CLG from a modal perspective, to reason about players' strategic abilities in the learning/teaching game. As mentioned already, sabotage modal logic SML is a suitable tool to characterize the original game SG, which extends the basic modal logic with a sabotage modality $\blacklozenge\varphi$, stating that there is an edge such that, φ is true at the evaluation node after deleting the edge from the model. However, given the differences between SG and CLG, we are going to develop a richer *modal logic of correct learning* CLL to capture the CLG framework.

Outline. Section 3.2 introduces CLL along with its application to CLG and some preliminary observations. Section 3.3 studies the expressivity of CLL. Section 3.4 investigates the model checking problem and satisfiability problem for CLL. We end this part by Section 3.5 on conclusion and future work.

3.2 A modal logic of correct learning

In this section, we introduce the language and semantics of CLL, and analyze its applications to CLG. Also, we make various observations, including some logical validities and relations between CLL and other existing logics.

3.2.1 Language and semantics

We begin by considering the action of Learner. In SML, the standard modality $\Diamond$ characterizes the transition from a node to its successors and corresponds well to Learner's actions in SG. However, operator $\Diamond$ is not any longer sufficient in our case. Note that after Teacher cuts a link $\langle w, v\rangle$ from Learner's current process S, Learner should start from w with the new path $S|_{\langle w,v\rangle}$ in the next round. Therefore, the desired operator should remember the history of Learner's movements, similar to the case of memory logics.

To capture Teacher's action, a natural place to start is by defining operators corresponding to link addition and deletion. There is already a body of literature on logics of these modalities, such as the sabotage operator $\blacklozenge$ and the *bridge operator* (Areces et al., 2012, 2015, 2018). As mentioned, each occurrence of $\blacklozenge$ in a formula deletes exactly one link, whereas the bridge operator *adds* links stepwise to models. Yet, including these two modalities is still not enough: we need to take into account if or not a link deleted by Teacher occurs in the path of Learner's movements. We now introduce the language $\mathcal{L}_s$ of CLL.

Definition 3.2: Let $\mathbf{P}$ be a countable set of propositional atoms. The *language* $\mathcal{L}_s$ is

recursively defined in the following way:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \diamondsuit\varphi \mid \langle - \rangle_{on}\varphi \mid \langle - \rangle_{off}\varphi \mid \langle + \rangle\varphi$$

where $p \in \mathbf{P}$. Notions $\top$, $\bot$, $\vee$ and $\rightarrow$ are as usual. Also, we use $\boxdot, [-]_{on}, [-]_{off}$ and $[+]$ to denote the dual operators of $\diamondsuit$, $\langle - \rangle_{on}$, $\langle - \rangle_{off}$ and $\langle + \rangle$ respectively.

Intuitively, $\diamondsuit\varphi$ states that φ holds after extending the current path with one of its successors. $\langle - \rangle_{on}\varphi$ reads φ is the case after deleting a link *on* the current path, while $\langle - \rangle_{off}\varphi$ states that after removing a link that is not on the path, φ holds. We use different subscripts '*on*' and '*off*' to indicate the two situations. Instead of link deletion, $\langle + \rangle\varphi$ shows that after extending the model with a particular link, φ holds. Roughly, operator $\diamondsuit$ is used to capture the actions of Learner in CLG, and operators $\langle + \rangle$, $\langle - \rangle_{on}$ and $\langle - \rangle_{off}$ characterize those of Teacher. This will become clear after we introduce the semantics.

Several fragments of $\mathcal{L}_s$ will be studied in the chapter. For brevity, we use a notational convention listing in subscript all modalities of the corresponding language. For instance, $\mathcal{L}_{\diamondsuit}$ is the fragment of $\mathcal{L}_s$ that has only the operator $\diamondsuit$ (besides Boolean connectives $\neg$ and $\wedge$); $\mathcal{L}_{\langle - \rangle_{off}}$ has only the modality $\langle - \rangle_{off}$; $\mathcal{L}_{\diamondsuit\langle - \rangle_{on}}$ has only $\diamondsuit$ and $\langle - \rangle_{on}$, etc. We now proceed to define the models.

Definition 3.3: A *model* of CLL is a tuple $\mathcal{M} = \langle W, R_L, R_T, V \rangle$, where W is a non-empty set of possible worlds, $R_{i \in \{L,T\}} \subseteq W \times W$ are two binary relations and $V : \mathbf{P} \rightarrow \mathcal{P}(W)$ is a valuation function. $\mathcal{F} = \langle W, R_L, R_T \rangle$ is a *frame*. Let S be an R_L-sequence, i.e., $Set(S) \subseteq R_L$. We name $\langle \mathcal{M}, S \rangle$ a *pointed model*, and S an *evaluation sequence*.

For brevity, usually we write $\mathcal{M}, S$ instead of $\langle \mathcal{M}, S \rangle$. Also, we use $\mathfrak{M}$ to denote the class of pointed models and $\mathfrak{M}^{\bullet}$ the class of pointed models whose sequence S is a singleton. Let $\mathcal{M} = \langle W, R_L, R_T, V \rangle$ be a model, $w \in W$ and $i \in \{L, T\}$. We use $R_i(w) := \{v \in W \mid R_i wv\}$ to denote the set of R_i-successors of w in $\mathcal{M}$. For any sequence S, define $R_i(S) := R_i(e(S))$, i.e., the R_i-successors of a sequence S are exactly the R_i-successors of its last element. Moreover, $\mathcal{M} \ominus \langle u, v \rangle := \langle W, R_L \setminus \{\langle u, v \rangle\}, R_T, V \rangle$ is the model obtained by removing $\langle u, v \rangle$ from R_L, and $\mathcal{M} \oplus \langle u, v \rangle := \langle W, R_L \cup \{\langle u, v \rangle\}, R_T, V \rangle$ is obtained by extending R_L in $\mathcal{M}$ with $\langle u, v \rangle$. Now let us introduce the semantics of CLL.

Definition 3.4: Let $\langle \mathcal{M}, S \rangle$ be a pointed model and $\varphi \in \mathcal{L}_s$. The *semantics* of CLL is defined as follows:

$$
\begin{aligned}
\mathcal{M}, S \vDash p &\text{ iff } e(S) \in V(p)\\
\mathcal{M}, S \vDash \neg\varphi &\text{ iff } \mathcal{M}, S \nvDash \varphi\\
\mathcal{M}, S \vDash \varphi \wedge \psi &\text{ iff } \mathcal{M}, S \vDash \varphi \text{ and } \mathcal{M}, S \vDash \psi\\
\mathcal{M}, S \vDash \diamondsuit\varphi &\text{ iff } \exists v \in R_L(S) \text{ s.t. } \mathcal{M}, S; v \vDash \varphi\\
\mathcal{M}, S \vDash \langle - \rangle_{on}\varphi &\text{ iff } \exists \langle v, v' \rangle \in Set(S) \setminus R_T \text{ s.t. } \mathcal{M} \ominus \langle v, v' \rangle, S|_{\langle v,v' \rangle} \vDash \varphi\\
\mathcal{M}, S \vDash \langle - \rangle_{off}\varphi &\text{ iff } \exists \langle v, v' \rangle \in (R_L \setminus R_T) \setminus Set(S) \text{ s.t. } \mathcal{M} \ominus \langle v, v' \rangle, S \vDash \varphi\\
\mathcal{M}, S \vDash \langle + \rangle\varphi &\text{ iff } \exists \langle v, v' \rangle \in R_T \setminus R_L \text{ s.t. } \mathcal{M} \oplus \langle v, v' \rangle, S \vDash \varphi
\end{aligned}
$$

Therefore, a propositional atom p is true at a sequence S iff the last element of S is p. Also, formula $\diamondsuit\varphi$ states that S has an R_L-successor v s.t. φ is true at $S; v$. Besides, $\langle - \rangle_{on}\varphi$ means that after deleting a link $\langle v, v' \rangle$ from $Set(S) \setminus R_T$, φ is true at $S|_{\langle v,v' \rangle}$. Moreover, $\langle - \rangle_{off}\varphi$ states that φ holds at S after cutting some link from $(R_L \setminus R_T) \setminus Set(S)$. Both the conditions for $\langle - \rangle_{on}$ and $\langle - \rangle_{off}$ require that the link deleted cannot be an R_T-edge. Intuitively, whereas $\langle - \rangle_{on}$ depicts the case when Teacher deletes a link from Learner's path S, $\langle - \rangle_{off}$ captures the situation that the link deleted does not occur in S. Finally, $\langle + \rangle\varphi$ means that after extending R_L with a new link of R_T, φ holds at the current sequence.

A formula φ is *satisfiable* if there exists $\langle \mathcal{M}, S \rangle \in \mathfrak{M}$ with $\mathcal{M}, S \vDash \varphi$. Also, *validity* in a model and in a frame is defined as usual. Note that the relevant class of pointed models to specify CLL is $\mathfrak{M}^{\bullet}$. Hence, CLL is the set of $\mathcal{L}_s$-formulas that are valid w.r.t. $\mathfrak{M}^{\bullet}$.

For any $\langle \mathcal{M}, S \rangle$ and $\langle \mathcal{M}', S' \rangle$, we say that they are *learning modal equivalent* (notation: $\langle \mathcal{M}, S \rangle \leftrightsquigarrow_l \langle \mathcal{M}', S' \rangle$) iff $\mathcal{M}, S \vDash \varphi \Leftrightarrow \mathcal{M}', S' \vDash \varphi$ for any $\varphi \in \mathcal{L}_s$. Besides, $\mathbb{T}^l(\mathcal{M}, S) := \{\varphi \in \mathcal{L}_s \mid \mathcal{M}, S \vDash \varphi\}$ denotes the CLL *theory* of S in $\mathcal{M}$. It is easy to see that two pointed models are learning modal equivalent if, and only if, they have the same CLL theory. In addition, we define a relation $\mathbf{U} \subseteq \mathfrak{M} \times \mathfrak{M}$ with $\langle \langle \mathcal{M}, S \rangle, \langle \mathcal{M}', S' \rangle \rangle \in \mathbf{U}$ iff $\langle \mathcal{M}', S' \rangle$ is $\langle \mathcal{M}, S; w \rangle$ for some state w with $\langle e(S), w \rangle \in R_L$, $\langle \mathcal{M} \ominus \langle v, v' \rangle, S|_{\langle v,v' \rangle} \rangle$ for some $\langle v, v' \rangle \in Set(S) \setminus R_T$, $\langle \mathcal{M} \ominus \langle v, v' \rangle, S \rangle$ for some $\langle v, v' \rangle \in (R_L \setminus R_T) \setminus Set(S)$, or $\langle \mathcal{M} \oplus \langle v, v' \rangle, S \rangle$ for some $\langle v, v' \rangle \in R_T \setminus R_L$. We can also iterate this order, to talk about models reachable in finitely many $\mathbf{U}$-steps, obtaining the relation $\mathbf{U}^*$.

3.2.2 Application: winning strategies in CLG

From Definition 3.4, it is easy to know that language $\mathcal{L}_s$ is able to capture the actions of both players in CLG. Also, our logic is expressive enough to describe the winning strategy (if there is one) for players in finite graphs.[1]

1 Generally speaking, to define the existence of winning strategies for players, we need to extend CLG with some fixpoint operators. We leave this for future inquiry.

Given a finite CLG, let p be a distinguished atom holding only at the goal node. Generally, the winning strategy of Learner and Teacher can be described by formulas of the following form:

$$\boxdot \bigcirc_0 \boxdot \bigcirc_1 \boxdot \cdots \bigcirc_n \boxdot (p \wedge [-]_{on}\bot) \tag{3-1}$$

where $\bigcirc_i$ is blank or one of $\langle - \rangle_{on}$, $\langle - \rangle_{\mathit{off}}$ and $\langle + \rangle$, for each $i \leqslant n \in \mathbb{N}$. In formula (3-1), the recurring $\boxdot$ operator depicts Learner's actions and $\bigcirc_i$ Teacher's response. The proposition p signalizes Learner's arrival at the goal, and $[-]_{on}\bot$ states that there are no edges in Learner's path that Teacher can cut. Hence, we conclude that Learner has reached the goal in a coherent way. Recall the example of CLG in Figure 3.1. Formula $\boxdot\langle + \rangle \boxdot \langle + \rangle \boxdot \langle - \rangle_{on} \boxdot \langle - \rangle_{\mathit{off}} \boxdot \boxdot (p \wedge [-]_{on}\bot)$ holds at the starting node a, so there exists a winning strategy in this specific CLG.

It is worthwhile to emphasis that in formula (3-1) we use $\boxdot$, other than $\diamondsuit$, to characterize the actions of Learner, which may be different from some other cases.[1] However, the modality $\boxdot$ used in formula (3-1) does not indicate that Learner is unwilling to learn. Essentially, it illustrates that she has no idea where to move in the next step, and we would claim that the form is in line with the spirit of CLG where Learner may move in wrong directions: Learner cannot distinguish different ways to the goal. In effect, all Learner can do in a CLG is to move as much as possible. Meanwhile, Teacher has to make some correct inferences 'visible' to Learner, and put Learner on track no matter what happens. Therefore, the form of formula (3-1) does not violate the cooperative nature of Learner.[2]

Remark 3.2: In SG we know that links cut by Teacher represent wrong inferences. However, SG does not tell us anything about the links that remain in the graph. Therefore, winning strategies of the players in SG cannot guarantee against situations like Gettier cases. In contrast, the formula $[-]_{on}\bot$ in formula (3-1) ensures that Teacher is not allowed to remove any more links from Learner's path. In CLG, a Gettier-style case is that Learner arrives at the goal node with some $\langle u, v \rangle \in R_L \setminus R_T$ occurring in her path, so Teacher now would be able to cut those links. Therefore Gettier cases cannot be winning strategies in correct learning games.

1 For instance, in sabotage games, we use $\diamond$ to capture actions of Learner in formulas describing winning strategies if they exist (see Gierasimczuk et al., 2009).

2 In contrast, one extreme case of non-cooperative variants of CLG might be that Learner is allowed to stay at her current position in each round: she makes no efforts to reach the goal node.

3.2.3 Preliminary observations

In this section, we make some preliminary observations on CLL. In particular, we discuss the relations between CLL and other related logics, present some logical validities, and study some basic features of CLL. Let us begin with the relation between $\mathcal{L}_{\diamondsuit}$ and the standard modal logic.

Proposition 3.1: Let $\mathcal{M} = \langle W, R_L, R_T, V\rangle$ be a model. For any $\langle \mathcal{M}, S\rangle \in \mathfrak{M}$ and $\varphi \in \mathcal{L}_{\diamondsuit}$, it holds that

$$\mathcal{M}, S \vDash \varphi \Leftrightarrow \langle W, R_L, V\rangle, e(S) \vDash \varphi^*$$

where $\varphi^* \in \mathcal{L}_{\Box}$ is a standard modal formula obtained by replacing every occurrence of $\diamondsuit$ in φ with $\Diamond$.

Proof The proof is done by induction on the syntax of $\varphi \in \mathcal{L}_{\diamondsuit}$. The Boolean cases are trivial. When φ is $\diamondsuit\psi$, it holds that:

$$\begin{aligned}
\mathcal{M}, S \vDash \varphi \quad &\Leftrightarrow \quad \text{there exists } v \in R_L(e(S)) \text{ such that } \mathcal{M}, S; v \vDash \psi \\
&\Leftrightarrow \quad \text{there exists } v \in R_L(e(S)) \text{ such that } \langle W, R_L, V\rangle, v \vDash \psi^* \\
&\Leftrightarrow \quad \langle W, R_L, V\rangle, e(S) \vDash \varphi^*
\end{aligned}$$

The first equivalence follows from Definition 3.4 directly. By the inductive hypothesis, the second one holds. The last one holds by the semantics of the standard modal logic. ∎

Therefore, essentially the fragment $\mathcal{L}_{\diamondsuit}$ of $\mathcal{L}_s$ is the standard modal logic. Moreover, the operator $\langle - \rangle_{off}$ is much similar to the sabotage operator $\blacklozenge$:

Proposition 3.2: Let $\mathcal{M} = \langle W, R_L, R_T, V\rangle$ be a model, and $R = R_L \setminus R_T$. For any $\langle \mathcal{M}, w\rangle \in \mathfrak{M}^{\bullet}$ and $\varphi \in \mathcal{L}_{\langle - \rangle_{off}}$, we have

$$\mathcal{M}, w \vDash \varphi \Leftrightarrow \langle W, R, V\rangle, w \vDash \varphi'$$

where $\varphi' \in \mathcal{L}_{\blacklozenge}$ is a SML formula obtained by replacing each occurrence of $\langle - \rangle_{off}$ in φ with $\blacklozenge$.

Proof It goes by induction on the structure of $\varphi \in \mathcal{L}_{\langle - \rangle_{off}}$. The Boolean cases are straightforward. When φ is $\langle - \rangle_{off}\psi$, it holds that:

$$\begin{aligned}
\mathcal{M}, w \vDash \varphi \quad &\Leftrightarrow \quad \text{there exists } \langle v, v'\rangle \in (R_L \setminus R_T) \text{ such that } \mathcal{M} \ominus \langle v, v'\rangle, w \vDash \psi \\
&\Leftrightarrow \quad \text{there exists } \langle v, v'\rangle \in R \text{ such that } \langle W, R \setminus \{\langle v, v'\rangle\}, V\rangle, w \vDash \psi' \\
&\Leftrightarrow \quad \langle W, R, V\rangle, w \vDash \blacklozenge\psi'
\end{aligned}$$

This completes the proof. ∎

Next, the following result captures the relation between $\mathcal{L}_{\diamondsuit\langle+\rangle}$ and the '*bridge modal logic* BML' (i.e., the modal logic extending the standard modal logic with the bridge operator):

Proposition 3.3: Let $\mathcal{M} = \langle W, R_L, W \times W, V\rangle$ be a model. For any $\langle \mathcal{M}, S\rangle \in \mathfrak{M}$ and $\varphi \in \mathcal{L}_{\diamondsuit\langle+\rangle}$, it holds that

$$\mathcal{M}, S \vDash \varphi \Leftrightarrow \langle W, R_L, V\rangle, e(S) \vDash \varphi^\star$$

where $\varphi^\star$ is a BML formula obtained by replacing every occurrence of $\diamondsuit$ in φ with $\Diamond$.[1]

Proof This goes by induction on the syntax of φ. The Boolean cases are trivial. The case for $\diamondsuit$ is similar to that of the proof of Proposition 3.1. When φ is $\langle+\rangle\psi$, it holds that:

$$\begin{aligned}
\mathcal{M}, S \vDash \varphi \quad &\Leftrightarrow \quad \exists\langle v, v'\rangle \in (R_T \setminus R_L) \text{ s.t. } \mathcal{M} \oplus \langle v, v'\rangle, S \vDash \psi \\
&\Leftrightarrow \quad \exists v, v' \in W \text{ s.t. } \langle v, v'\rangle \notin R_L \text{ and } \langle W, R_L \cup \{\langle v, v'\rangle\}, V\rangle, e(S) \vDash \psi^\star \\
&\Leftrightarrow \quad \langle W, R_L, V\rangle, e(S) \vDash \langle+\rangle\psi^\star
\end{aligned}$$

The proof is completed. ∎

From Proposition 3.1-3.3, we know that several fragments of CLL are similar to some existing logics. Yet, as a whole, different operators of CLL interact with each other. For instance, for any $\langle \mathcal{M}, w\rangle \in \mathfrak{M}^\bullet$, formula $[-]_{on}\varphi$ is valid, as $Set(w) = \varnothing$. However, $\diamondsuit\neg[-]_{on}\varphi$ is satisfiable. This presents a drastic difference between CLL and other logics mentioned so far: in those logics, it is impossible that the evaluation point has access to a node satisfying a contradiction. To understand how operators in CLL work, we present some other validities.

Proposition 3.4: Let $p \in \mathbf{P}$ and $\varphi, \psi \in \mathcal{L}_s$. All the following formulas are validities of logic CLL (w.r.t. $\mathfrak{M}^\bullet$):

$$p \wedge \diamondsuit\top \to \boxdot[-]_{on}p \tag{3-2}$$

$$\bigcirc(\varphi \to \psi) \to (\bigcirc\varphi \to \bigcirc\psi) \qquad \bigcirc \in \{[-]_{off}, [+]\} \tag{3-3}$$

$$\boxdot^n [-]_{on}(\varphi \to \psi) \to (\boxdot^n[-]_{on}\varphi \to \boxdot^n[-]_{on}\psi) \qquad n \in \mathbb{N} \tag{3-4}$$

$$\diamondsuit^n\langle-\rangle_{on}\varphi \to \bigvee_{m<n} \diamondsuit^m\langle-\rangle_{off}\varphi \qquad 1 \leqslant n \in \mathbb{N} \tag{3-5}$$

1 By abuse of notation, for any $\varphi \in \mathcal{L}_{\diamondsuit\langle+\rangle}$, $\varphi^\star$ is a formula of the bridge modal logic.

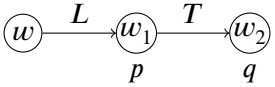


Figure 3.2 A case showing that validities of $\mathcal{L}_{\diamondsuit\langle-\rangle_{on}}$ are not closed under substitution. Consider the general schema $\varphi \wedge \diamondsuit\psi \rightarrow \boxdot[-]_{on}\varphi$ of formula (3-2). Let $\varphi := \diamondsuit p$ and $\psi := \boxdot q$. It holds that $\mathcal{M}, w \vDash \diamondsuit p \wedge \diamondsuit \boxdot q$. But, since w has exactly one R_L-successor w_1 and $\langle w, w_1\rangle \notin R_T$, we have $\mathcal{M}, w \nvDash \boxdot[-]_{on}\diamondsuit p$.

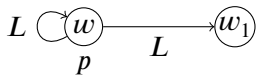


Figure 3.3 A model of $\varphi_{\mathbb{T}}$. It is not hard to see that $\varphi_{\mathbb{T}}$ is true at w.

Their validity holds immediately by the semantics. Formula (3-2) states that, for any singleton w, if it is p and has some R_L-successors, then any of its extensions $\langle w, v\rangle$ with $v \in R_L(w)$ is $[-]_{on}p$, no matter whether $\langle w, v\rangle \in R_T$ or not. Principles (3-3) and (3-4) show that all operators $[-]_{off}$, $[+]$ and $[-]_{on}$ are normal operators. Formula (3-5) illustrates that in some situations, a formula containing $\langle-\rangle_{on}$ can be reduced to another formula containing $\langle-\rangle_{off}$.

Note that principle (3-2) is not a schema. Although it will still be valid if we replace propositional atoms occurring in it with any other Boolean formulas, substitution fails generally. See Figure 3.2 for an example, which essentially illustrates the following result:

Proposition 3.5: $\mathcal{L}_{\diamondsuit\langle-\rangle_{on}}$ and CLL are not closed under substitution.

Moreover, CLL and $\mathcal{L}_{\diamondsuit\langle-\rangle_{on}}$ also have other features very different from the standard modal logic. For instance,

Proposition 3.6: Both $\mathcal{L}_{\diamondsuit\langle-\rangle_{on}}$ and CLL lack the tree model property.

Proof Let $\varphi_{\mathbb{T}}$ be the conjunction of the following:

$$(\mathbb{T}_1) \qquad p \wedge \diamondsuit p \wedge \diamondsuit\neg p$$

$$(\mathbb{T}_2) \qquad \boxdot(p \rightarrow \diamondsuit p \wedge \diamondsuit\neg p)$$

$$(\mathbb{T}_3) \qquad \boxdot(\neg p \rightarrow \langle-\rangle_{on}(\boxdot p \wedge \boxdot\boxdot p))$$

It is not hard to see that formula $\varphi_{\mathbb{T}} \in \mathcal{L}_{\diamondsuit\langle-\rangle_{on}}$ is satisfiable w.r.t. $\mathfrak{M}^{\bullet}$ (see Figure 3.3). We now show that, for any $\mathcal{M} = \langle W, R_L, R_T, V\rangle$ and $w \in W$, $\mathcal{M}, w \vDash \varphi_{\mathbb{T}}$ entails $R_L ww$. By $(\mathbb{T}_1)$, it follows that $w \in V(p)$, and it can reach some $w_1 \in V(p)$ and some $w_2 \notin V(p)$ via R_L. Besides, $(\mathbb{T}_2)$ states that, via R_L, each such w_1 can also reach some p-node w_3 and $\neg p$-node w_4. Finally, from $(\mathbb{T}_3)$ we know that w can only reach one $\neg p$-point

by R_L and that w_1 does not have $\neg p$-successors via R_L any longer after cutting $\langle w, w_2\rangle$. So, $\langle w, w_2\rangle = \langle w_1, w_4\rangle$. Therefore, $R_L ww$. The proof is completed. ∎

As observed, CLL has some distinguishing features. In the sections to come we will make a deeper investigation into our logic.

3.3 Expressive power of CLL

In this section, we study the expressivity of CLL. First, we will show that CLL is still a fragment of FOL even though it looks complicated. After this, a suitable notion of bisimulation for CLL is introduced. Finally, we provide a van Benthem style characterization theorem for the logic.

3.3.1 First-order translation

Given the complicated semantics, is CLL still a fragment of FOL? In this part we will provide a positive answer to this question, by describing a translation from CLL to FOL.

It is not hard to see that the first-order language here cannot be $\mathcal{L}_1$ used to translate logic $\mathsf{S_d ML}$ given in Section 2.3, as we now have two relations. Let $\mathcal{L}_1^{\dagger}$ be the first-order language consisting of countable unary predicates $P_{i\in\mathbb{N}}$, two binary relations $R_{i\in\{L,T\}}$, and equality $\equiv$. Take any finite, non-empty sequence E of variables. Let y, y' be two fresh variables not appearing in E. When there exists $\langle x, x'\rangle \in Set(E)$ with $x \equiv y$ and $x' \equiv y'$, we define $E|_{\langle y,y'\rangle} := E|_{\langle x,x'\rangle}$. Now let us define the first-order translation.

Definition 3.5: Let $E = \langle x_0, x_1, \cdots, x_n\rangle$ be a finite sequence (non-empty) of variables without any variable occurring more than once, and E^- and E^+ two finite sets (maybe empty) of ordered pairs of variables. The *first-order translation* $\mathfrak{T}(\varphi, E, E^+, E^-)$ from $\varphi \in \mathcal{L}_s$ to first-order formulas is as follows:

$$
\begin{aligned}
\mathfrak{T}(p, E, E^+, E^-) &= Pe(E)\\
\mathfrak{T}(\neg\varphi, E, E^+, E^-) &= \neg\mathfrak{T}(\varphi, E, E^+, E^-)\\
\mathfrak{T}(\varphi\wedge\psi, E, E^+, E^-) &= \mathfrak{T}(\varphi, E, E^+, E^-)\wedge\mathfrak{T}(\psi, E, E^+, E^-)\\
\mathfrak{T}(\dot{\Diamond}\varphi, E, E^+, E^-) &= \exists y\Big(\big(\bigvee_{\langle x,x'\rangle\in E^+}(e(E)\equiv x\wedge y\equiv x')\vee(R_L e(E) y\wedge\\
&\qquad \neg\bigvee_{\langle v,v'\rangle\in E^-}(e(E)\equiv v\wedge y\equiv v'))\big)\wedge\mathfrak{T}(\varphi, E; y, E^+, E^-)\Big)
\end{aligned}
$$

$$\mathfrak{T}(\langle - \rangle_{on}\varphi, E, E^+, E^-) = \exists y \exists y' \Big(\bigvee_{\langle x,x' \rangle \in Set(E) \setminus (E^- \cup E^+)} (y \equiv x \wedge y' \equiv x') \wedge R_L yy' \wedge \neg R_T yy' \wedge \mathfrak{T}(\varphi, E|_{\langle y,y' \rangle}, E^+, E^- \cup \{\langle y, y' \rangle\}) \Big)$$

$$\mathfrak{T}(\langle - \rangle_{off}\varphi, E, E^+, E^-) = \exists y \exists y' \Big(\neg \bigvee_{\langle x,x' \rangle \in Set(E) \cup E^- \cup E^+} (y \equiv x \wedge y' \equiv x') \wedge R_L yy' \wedge \neg R_T yy' \wedge \mathfrak{T}(\varphi, E, E^+, E^- \cup \{\langle y, y' \rangle\}) \Big)$$

$$\mathfrak{T}(\langle + \rangle\varphi, E, E^+, E^-) = \exists y \exists y' \Big(\neg \bigvee_{\langle x,x' \rangle \in E^- \cup E^+} (y \equiv x \wedge y' \equiv x') \wedge \neg R_L yy' \wedge R_T yy' \wedge \mathfrak{T}(\varphi, E, E^+ \cup \{\langle y, y' \rangle\}, E^-) \Big)$$

where y, y' are variables having not been used yet in the translation. In addition, given a set Φ of $\mathcal{L}_s$-formulas, we denote by $\mathfrak{T}(\Phi, E, E^+, E^-)$ the set $\{\mathfrak{T}(\varphi, E, E^+, E^-) \mid \varphi \in \Phi\}$ of first-order translations of formulas in Φ.

From the perspective of correct learning games, sequence E stands for Learner's current process, and E^+, E^- represent links having already been added and deleted respectively. In any translation, both sets E^+ and E^- may be extended. For any their extensions $E^+ \cup X$ and $E^- \cup Y$, it holds that $X \cap Y = \varnothing$. This is in line with our semantics: links deleted are different from those added. Furthermore, unlike the standard modal logic, generally the translation does not yield a formula with only one free variable. But, it does so when setting E, E^+ and E^- to be a singleton, $\varnothing$ and $\varnothing$ respectively.

Note that in Definition 3.5, the sequence E includes no variable appearing more than once, and it is easy to see that any modification of E in a translation still has this property. Specifically, this requirement is used to guarantee that assignments are well-defined. Let σ be an assignment, S a sequence of points in a model, and E a sequence of variables with the same size as S. In what follows, when writing $\sigma_{E:=S}$, we mean a new assignment that is the same as σ except assigning variables in E to the corresponding elements in S. Since all variables in E appear only once, no variable in the sequence can be assigned to different elements in S. With Definition 3.5, we have the following result:

Lemma 3.1: Let $\mathfrak{T}(\varphi, E, E^+, E^-)$ be a translation with $E^+ \cap E^- = \varnothing$, and y, y' two fresh variables. For any σ and $\mathcal{M}$, it holds that $\mathcal{M} \ominus \langle v, v' \rangle \vDash \mathfrak{T}(\varphi, E, E^+, E^-)[\sigma]$ iff $\mathcal{M} \vDash \mathfrak{T}(\varphi, E, E^+, E^- \cup \{\langle y, y' \rangle\})[\sigma_{y^{(\prime)}:=v^{(\prime)}}]$, for any $\langle v, v' \rangle \in R_L \setminus R_T$; and $\mathcal{M} \oplus \langle v, v' \rangle \vDash \mathfrak{T}(\varphi, E, E^+, E^-)[\sigma]$ iff $\mathcal{M} \vDash \mathfrak{T}(\varphi, E, E^+ \cup \{\langle y, y' \rangle\}, E^-)[\sigma_{y^{(\prime)}:=v^{(\prime)}}]$, for any $\langle v, v' \rangle \in R_T \setminus R_L$.

Proof The proofs for these two cases are similar, and both of them can be shown by induction on the syntax of formulas. We focus on the first one, and only prove the cases for propositional atoms and $\langle - \rangle_{on}$. Assume that $\langle v, v' \rangle \in R_L \setminus R_T$, and $R_L^- := R_L \setminus \{\langle v, v' \rangle\}$.

(1). Formula φ is $p \in \mathbf{P}$. By Definition 3.5, $\mathcal{M} \ominus \langle v, v' \rangle \vDash \mathfrak{T}(\varphi, E, E^+, E^-)[\sigma]$ iff $\mathcal{M} \ominus \langle v, v' \rangle \vDash Pe(E)[\sigma]$. From the definition of $\mathcal{M} \ominus \langle v, v' \rangle$, it follows that $\mathcal{M} \ominus \langle v, v' \rangle \vDash Pe(E)[\sigma]$ iff $\mathcal{M} \vDash Pe(E)[\sigma]$. Again, by Definition 3.5, it holds that $\mathcal{M} \vDash Pe(E)[\sigma]$ iff $\mathcal{M} \vDash \mathfrak{T}(\varphi, E, E^+, E^- \cup \{\langle y, y' \rangle\})[\sigma_{y^{(\prime)} := v^{(\prime)}}]$.

(2). Formula φ is $\langle - \rangle_{on} \psi$. We have the following equivalences:

$$
\begin{aligned}
& \mathcal{M} \ominus \langle v, v' \rangle \vDash \mathfrak{T}(\varphi, E, E^+, E^-)[\sigma] \\
\Leftrightarrow\ & \mathcal{M} \ominus \langle v, v' \rangle \vDash \exists u \exists u' \Big(\bigvee_{\langle z, z' \rangle \in Set(E) \setminus (E^- \cup E^+)} (u \equiv z \wedge u' \equiv z') \wedge R_L^- u u' \wedge \\
& \qquad \neg R_T u u' \wedge \mathfrak{T}(\psi, E|_{\langle u, u' \rangle}, E^+, E^- \cup \{\langle u, u' \rangle\}) \Big)[\sigma] \\
\Leftrightarrow\ & \mathcal{M} \vDash \exists u \exists u' \Big(\bigvee_{\langle z, z' \rangle \in Set(E) \setminus (E^+ \cup E^- \cup \{\langle y, y' \rangle\})} (u \equiv z \wedge u' \equiv z') \wedge R_L u u' \wedge \\
& \qquad \neg R_T u u' \wedge \mathfrak{T}(\psi, E|_{\langle u, u' \rangle}, E^+, E^- \cup \{\langle u, u' \rangle, \langle y, y' \rangle\}) \Big)[\sigma_{y^{(\prime)} := v^{(\prime)}}] \\
\Leftrightarrow\ & \mathcal{M} \vDash \mathfrak{T}(\varphi, E, E^+, E^- \cup \{\langle y, y' \rangle\})[\sigma_{y^{(\prime)} := v^{(\prime)}}]
\end{aligned}
$$

The first equivalence holds directly by Definition 3.5. By the inductive hypothesis and the definition of R_L^-, the second one holds. The last equivalence follows from the definition of first-order translation. The proof is completed. ∎

With Lemma 3.1, we now can show the correctness of the translation:

Theorem 3.1: Let $\langle \mathcal{M}, S \rangle \in \mathfrak{M}$ and E an R_L-sequence of variables with the same size as S. For any $\varphi \in \mathcal{L}_s$, $\mathcal{M}, S \vDash \varphi$ iff $\mathcal{M} \vDash \mathfrak{T}(\varphi, E, \emptyset, \emptyset)[\sigma_{E := S}]$.

Proof The proof is by induction on the structure of $\varphi \in \mathcal{L}_s$. Also, we only consider the cases for propositional atoms and $\langle - \rangle_{on}$.

(1). Formula φ is $p \in \mathbf{P}$. By the semantics, $\mathcal{M}, S \vDash \varphi$ iff $e(S) \in V(p)$. On the other hand, by Definition 3.5, $\mathfrak{T}(\varphi, E, \emptyset, \emptyset)$ is $Pe(E)$. So we have $\mathcal{M}, S \vDash \varphi$ iff $\mathcal{M} \vDash \mathfrak{T}(\varphi, E, \emptyset, \emptyset)[\sigma_{E := S}]$.

(2). When φ is $\langle - \rangle_{on} \psi$, the following equivalences hold:

$$
\begin{aligned}
& \mathcal{M}, S \vDash \varphi \\
\Leftrightarrow\ & \text{there exists } \langle v, v' \rangle \in (Set(S) \setminus R_T) \text{ s.t. } \mathcal{M} \ominus \langle v, v' \rangle, S|_{\langle v,v' \rangle} \vDash \psi \\
\Leftrightarrow\ & \text{there exists } \langle v, v' \rangle \in (Set(S) \setminus R_T) \text{ s.t.} \\
& \mathcal{M} \ominus \langle v, v' \rangle \vDash \mathfrak{T}(\psi, E|_{\langle y,y' \rangle}, \varnothing, \varnothing)[\sigma_{E:=S, y^{(\prime)}:=v^{(\prime)}}] \\
\Leftrightarrow\ & \mathcal{M} \vDash \exists y \exists y' (\bigvee_{\langle v,v' \rangle \in Set(E)} (y \equiv v \wedge y' \equiv v') \wedge R_L y y' \wedge \neg R_T y y' \wedge \\
& \qquad \mathfrak{T}(\psi, E|_{\langle y,y' \rangle}, \varnothing, \{\langle y, y' \rangle\}))[\sigma_{E:=S}] \\
\Leftrightarrow\ & \mathcal{M} \vDash \mathfrak{T}(\varphi, E, \varnothing, \varnothing)[\sigma_{E:=S}]
\end{aligned}
$$

The first equivalence follows from our semantics immediately. By the inductive hypothesis, the second one follows. With Lemma 3.1, we have the third one. The last one follows directly from Definition 3.5. This completes the proof. ∎

In the result above, we have an extra requirement on the sequence E used in the translation, i.e., $Set(E) \subseteq R_L$. Intuitively, this restriction corresponds to the definition of pointed models. When S is a singleton, E is also a singleton, and each extension of E fulfils the requirement automatically by Definition 3.5.

So far, by the translation, we have shown that CLL is a fragment of FOL. Also, Definition 3.5 gives us other information about our logic. For example, it includes immediate transfer of the compactness property of FOL to CLL. Moreover, since the complexity of the model checking problem for FOL is PSPACE-complete and the translation has only a polynomial size increase, we can obtain an upper bound for that of CLL. We will return to this below.

3.3.2 Bisimulation and characterization for CLL

In this part, we continue to study the expressive power of CLL. In particular, we introduce a novel notion of 'learning bisimulation (l-bisimulation)' for our logic, which finally leads to a van Benthem style characterization theorem.

Definition 3.6: For any $\mathcal{M} = \langle W, R_L, R_T, V \rangle$ and $\mathcal{M}' = \langle W', R'_L, R'_T, V' \rangle$, a non-empty relation $Z_l \subseteq \mathbf{U}^*(\langle \mathcal{M}, S \rangle) \times \mathbf{U}^*(\langle \mathcal{M}', S' \rangle)$ is an *l-bisimulation* between $\langle \mathcal{M}, S \rangle$ and $\langle \mathcal{M}', S' \rangle$ (notation: $\langle \mathcal{M}, S \rangle Z_l \langle \mathcal{M}', S' \rangle$) if:

Atom: $\mathcal{M}, S \vDash p$ iff $\mathcal{M}', S' \vDash p$, for each $p \in \mathbf{P}$.

Zig$_{\Diamond}$: If there exists $v \in R_L(e(S))$, then there exists $v' \in R'_L(e(S'))$ such that $\langle \mathcal{M}, S; v \rangle Z_l \langle \mathcal{M}', S'; v' \rangle$.

Zig$_{\langle - \rangle_{on}}$: If there is $\langle u, v \rangle \in Set(S) \setminus R_T$, then there is $\langle u', v' \rangle \in Set(S') \setminus R'_T$ with $\langle \mathcal{M} \ominus \langle u, v \rangle, S|_{\langle u,v \rangle} \rangle Z_l \langle \mathcal{M}' \ominus \langle u', v' \rangle, S'|_{\langle u',v' \rangle} \rangle$.

Zig$_{\langle - \rangle_{off}}$**:** If there exists $\langle u, v\rangle \in (R_L \setminus R_T) \setminus Set(S)$, then there exists $\langle u', v'\rangle \in (R'_L \setminus R'_T) \setminus Set(S')$ with $\langle \mathcal{M} \ominus \langle u, v\rangle, S\rangle Z_l \langle \mathcal{M}' \ominus \langle u', v'\rangle, S'\rangle$.

Zig$_{\langle + \rangle}$**:** If there exists $\langle u, v\rangle \in R_T \setminus R_L$, then there exists $\langle u', v'\rangle \in R'_T \setminus R'_L$ with $\langle \mathcal{M} \oplus \langle u, v\rangle, S\rangle Z_l \langle \mathcal{M}' \oplus \langle u', v'\rangle, S'\rangle$.

Zag$_{\diamondsuit}$, **Zag**$_{\langle - \rangle_{on}}$, **Zag**$_{\langle - \rangle_{off}}$ and **Zag**$_{\langle + \rangle}$: the analogous clauses in the converse direction of **Zig**$_{\diamondsuit}$, **Zig**$_{\langle - \rangle_{on}}$, **Zig**$_{\langle - \rangle_{off}}$ and **Zig**$_{\langle + \rangle}$ respectively.

For brevity, we usually write $\langle \mathcal{M}, S\rangle \underline{\leftrightarrow}_l \langle \mathcal{M}', S'\rangle$ if there is an l-bisimulation Z_l such that $\langle \mathcal{M}, S\rangle Z_l \langle \mathcal{M}', S'\rangle$.

The clauses for $\diamondsuit\!\!\!\!\cdot$ is similar to those for $\diamondsuit$ in the standard bisimulation: they keep the model fixed and extend the evaluation sequence with one of its R_L-successors. However, all conditions for $\langle - \rangle_{on}$, $\langle - \rangle_{off}$ and $\langle + \rangle$ change the model. In particular, clauses for $\langle - \rangle_{off}$ and $\langle + \rangle$ do not modify the evaluation sequence, while those for $\langle - \rangle_{on}$ change both the model and the current sequence. By a straightforward induction on $\varphi \in \mathcal{L}$, we have the following result:

Theorem 3.2 ($\underline{\leftrightarrow}_l \subseteq \leftrightsquigarrow_l$): For any pointed models $\langle \mathcal{M}, S\rangle$ and $\langle \mathcal{M}', S'\rangle$, it holds that:

$$\langle \mathcal{M}, S\rangle \underline{\leftrightarrow}_l \langle \mathcal{M}', S'\rangle \Rightarrow \langle \mathcal{M}, S\rangle \leftrightsquigarrow_l \langle \mathcal{M}', S'\rangle.$$

Proof It goes by induction on φ. Assume that $\langle \mathcal{M}, S\rangle \underline{\leftrightarrow}_l \langle \mathcal{M}', S'\rangle$. The Boolean cases are straightforward.

(1). φ is $\diamondsuit\!\!\!\!\cdot\psi$. If $\mathcal{M}, S \vDash \varphi$, then there exists $v \in R_L(S)$ such that $\mathcal{M}, S; v \vDash \psi$. By **Zig**$_{\diamondsuit}$, there exists $v' \in R'_L(S')$ such that $\langle \mathcal{M}, S; v\rangle \underline{\leftrightarrow}_l \langle \mathcal{M}', S'; v'\rangle$. By the inductive hypothesis, it holds that $\langle \mathcal{M}, S; v\rangle \leftrightsquigarrow_l \langle \mathcal{M}', S'; v'\rangle$. Consequently, $\mathcal{M}', S'; v' \vDash \psi$, which is followed by $\mathcal{M}', S' \vDash \varphi$ immediately. Similarly, we can obtain $\mathcal{M}, S \vDash \varphi$ from $\mathcal{M}', S' \vDash \varphi$ by **Zag**$_{\diamondsuit}$.

(2). Formula φ is $\langle - \rangle_{on}\psi$. When $\mathcal{M}, S \vDash \varphi$, there exists $\langle u, v\rangle \in Set(S) \setminus R_T$ with $\mathcal{M} \ominus \langle u, v\rangle, S|_{\langle u,v\rangle} \vDash \psi$. By **Zig**$_{\langle - \rangle_{on}}$, there exists $\langle u', v'\rangle \in Set(S') \setminus R'_T$ such that $\langle \mathcal{M} \ominus \langle u, v\rangle, S|_{\langle u,v\rangle}\rangle \underline{\leftrightarrow}_l \langle \mathcal{M}' \ominus \langle u', v'\rangle, S'|_{\langle u',v'\rangle}\rangle$. By the inductive hypothesis, we have $\langle \mathcal{M} \ominus \langle u, v\rangle, S|_{\langle u,v\rangle}\rangle \leftrightsquigarrow_l \langle \mathcal{M}' \ominus \langle u', v'\rangle, S'|_{\langle u',v'\rangle}\rangle$. So, $\mathcal{M}' \ominus \langle u', v'\rangle, S'|_{\langle u',v'\rangle} \vDash \psi$. Now it follows that $\mathcal{M}', S' \vDash \varphi$. In a similar way, when $\mathcal{M}', S' \vDash \varphi$, we can prove $\mathcal{M}, S \vDash \varphi$ by **Zag**$_{\langle - \rangle_1}$.

(3). Formula φ is $\langle - \rangle_{off}\psi$. If $\mathcal{M}, S \vDash \varphi$, then there is $\langle u, v\rangle \in (R_L \setminus R_T) \setminus Set(S)$ with $\mathcal{M} \ominus \langle u, v\rangle, S \vDash \varphi$. By **Zig**$_{\langle - \rangle_{off}}$, there exists $\langle u', v'\rangle \in (R'_L \setminus R'_T) \setminus Set(S')$ such that $\langle \mathcal{M} \ominus \langle u, v\rangle, S\rangle \underline{\leftrightarrow}_l \langle \mathcal{M}' \ominus \langle u', v'\rangle, S'\rangle$. By the inductive hypothesis, it follows that

$\langle \mathcal{M} \ominus \langle u, v\rangle, S\rangle \leftrightsquigarrow_l \langle \mathcal{M}' \ominus \langle u', v'\rangle, S'\rangle$. Consequently, $\mathcal{M}' \ominus \langle u', v'\rangle, S' \vDash \varphi$. So we have $\mathcal{M}', S' \vDash \varphi$. Similarly, when $\mathcal{M}', S' \vDash \varphi$, we can prove $\mathcal{M}, S \vDash \varphi$ by $\mathbf{Zag}_{\langle - \rangle_{off}}$.

(4). Finally, let us consider the case that formula φ is $\langle + \rangle \psi$. When $\mathcal{M}, S \vDash \varphi$, there exists $\langle u, v\rangle \in R_T \setminus R_L$ with $\mathcal{M} \oplus \langle u, v\rangle, S \vDash \psi$. By $\mathbf{Zig}_{\langle + \rangle}$, there exists $\langle u', v'\rangle \in R'_T \setminus R'_L$ with $\langle \mathcal{M} \oplus \langle u, v\rangle, S\rangle \underline{\leftrightarrow}_l \langle \mathcal{M}' \oplus \langle u', v'\rangle, S'\rangle$. By the inductive hypothesis, it holds that $\langle \mathcal{M} \oplus \langle u, v\rangle, S\rangle \leftrightsquigarrow_l \langle \mathcal{M}' \oplus \langle u', v'\rangle, S'\rangle$. Therefore, we have $\mathcal{M}' \oplus \langle u', v'\rangle, S' \vDash \psi$. Consequently, $\mathcal{M}', S' \vDash \varphi$. Similarly, by $\mathbf{Zag}_{\langle + \rangle}$, we know $\mathcal{M}, S \vDash \varphi$ from $\mathcal{M}', S' \vDash \varphi$. This completes the proof. ∎

Moreover, the converse direction of Theorem 3.2 holds for the models that are ω-saturated. For each finite set Y, we denote the expansion of $\mathcal{L}_1^\dagger$ with a set Y of constants with $\mathcal{L}_1^{\dagger Y}$, and denote the expansion of $\mathcal{M}$ to $\mathcal{L}_1^{\dagger Y}$ with $\mathcal{M}^Y$. Let $\mathbf{x}$ be a finite tuple of variables. In this setting, we say a model $\mathcal{M} = \langle W, R_L, R_T, V\rangle$ of CLL is *ω-saturated* if, for every finite subset Y of W, the expansion $\mathcal{M}^Y$ realizes every set $\Gamma(\mathbf{x})$ of $\mathcal{L}_1^{\dagger Y}$-formulas whose finite subsets $\Gamma'(\mathbf{x})$ are all realized in $\mathcal{M}^Y$.

Theorem 3.3 ($\leftrightsquigarrow_l \subseteq \underline{\leftrightarrow}_l$): For any pointed models $\langle \mathcal{M}, S\rangle$ and $\langle \mathcal{M}', S'\rangle$ of CLL that are ω-saturated, it holds that:

$$\langle \mathcal{M}, S\rangle \leftrightsquigarrow_l \langle \mathcal{M}', S'\rangle \Rightarrow \langle \mathcal{M}, S\rangle \underline{\leftrightarrow}_l \langle \mathcal{M}', S'\rangle.$$

Proof To prove this, we show that $\leftrightsquigarrow_l$ itself is an l-bisimulation. Here we only prove the cases involving clauses $\mathbf{Zig}_{\diamondsuit}$ and $\mathbf{Zig}_{\langle - \rangle_{on}}$. Let E' be a sequence of variables over R'_L with the same size as S'.

(1). Let $v \in R_L(S)$. We are going to prove that there is some $v' \in R'_L(S')$ such that $\langle \mathcal{M}, S; v\rangle \leftrightsquigarrow_l \langle \mathcal{M}', S'; v'\rangle$. For any finite $\Gamma \subseteq \mathbb{T}^l(\mathcal{M}, S; v)$, we have:

$$\begin{aligned} \mathcal{M}, S \vDash \diamondsuit \bigwedge \Gamma &\Leftrightarrow \mathcal{M}', S' \vDash \diamondsuit \bigwedge \Gamma \\ &\Leftrightarrow \mathcal{M}' \vDash \mathfrak{T}(\diamondsuit \bigwedge \Gamma, E', \varnothing, \varnothing)[\sigma_{E' := S'}] \\ &\Leftrightarrow \mathcal{M}' \vDash \exists y (R'_L e(E') y \wedge \mathfrak{T}(\bigwedge \Gamma, E'; y, \varnothing, \varnothing))[\sigma_{E' := S'}] \end{aligned}$$

As the pointed model $\langle \mathcal{M}', S'\rangle$ is ω-saturated, there exists $y \in R'_L(E')$ such that $\mathcal{M}' \vDash \mathfrak{T}(\mathbb{T}^l(\mathcal{M}, S; v), E'; y, \varnothing, \varnothing)[\sigma_{E' := S'}]$. By Theorem 3.1, there is $v' \in R'_L(S')$ such that $\langle \mathcal{M}, S; v\rangle \leftrightsquigarrow_l \langle \mathcal{M}', S'; v'\rangle$. The proof of the $\mathbf{Zig}_{\diamondsuit}$ clause is completed.

(2). Let $\langle u, v\rangle \in Set(S) \setminus R_T$. We will show that there exists $\langle u', v'\rangle \in Set(S') \setminus R'_T$ such that $\langle \mathcal{M} \ominus \langle u, v\rangle, S|_{\langle u, v\rangle}\rangle \leftrightsquigarrow_l \langle \mathcal{M}' \ominus \langle u', v'\rangle, S'|_{\langle u', v'\rangle}\rangle$. Let Γ be a finite subset of $\mathbb{T}^l(\mathcal{M} \ominus \langle u, v\rangle, S|_{\langle u, v\rangle})$, then the following equivalences hold:

$$\begin{aligned}\mathcal{M}, S \vDash \langle - \rangle_{on} \bigwedge \Gamma &\Leftrightarrow \mathcal{M}', S' \vDash \langle - \rangle_{on} \bigwedge \Gamma \\ &\Leftrightarrow \mathcal{M}' \vDash \mathfrak{T}(\langle - \rangle_{on} \bigwedge \Gamma, E', \varnothing, \varnothing)[\sigma_{E':=S'}] \\ &\Leftrightarrow \mathcal{M}' \vDash \exists y \exists z (\bigvee_{\langle x, x' \rangle \in Set(E')} (y \equiv x \wedge z \equiv x') \wedge \neg R'_T yz \wedge \\ &\qquad \mathfrak{T}(\bigwedge \Gamma, E'|_{\langle y,z \rangle}, \varnothing, \{\langle y, z \rangle\}))[\sigma_{E':=S'}]\end{aligned}$$

Since $\langle \mathcal{M}', S' \rangle$ is ω-saturated, there are y, z such that $\langle y, z \rangle \in Set(E') \setminus R'_T$ and $\mathcal{M}' \vDash \mathfrak{T}(\mathbb{T}^l(\mathcal{M} \ominus \langle u, v \rangle, S|_{\langle u,v \rangle}), E'|_{\langle y,z \rangle}, \varnothing, \{\langle y, z \rangle\}))[\sigma_{E':=S'}]$. Without loss of generality, we assume $\sigma(y) = u'$ and $\sigma(z) = v'$. As $\langle u', v' \rangle \in Set(S') \setminus R'_T$, from Lemma 3.1 it follows that $\mathcal{M}' \ominus \langle u', v' \rangle \vDash \mathfrak{T}(\mathbb{T}^l(\mathcal{M} \ominus \langle u, v \rangle, S|_{\langle u,v \rangle}), E'|_{\langle y,z \rangle}, \varnothing, \varnothing))[\sigma_{E':=S'}]$. By Theorem 3.1, we have $\mathcal{M}' \ominus \langle u', v' \rangle, S'|_{\langle u',v' \rangle} \vDash \mathbb{T}^l(\mathcal{M} \ominus \langle u, v \rangle, S|_{\langle u,v \rangle})$. So, it holds that $\langle \mathcal{M} \ominus \langle u, v \rangle, S|_{\langle u,v \rangle} \rangle \leftrightsquigarrow_l \langle \mathcal{M}' \ominus \langle u', v' \rangle, S'|_{\langle u',v' \rangle} \rangle$. The proof of $\mathbf{Zig}_{\langle - \rangle_{on}}$ is completed.∎

Thus, we have established a precise match between learning modal equivalence and learning bisimulation for the ω-saturated models of CLL. Now, by a simple adaptation of standard arguments (Aucher et al., 2018; Blackburn et al., 2001; Li, 2020), we can show the following result:

Theorem 3.4: For any $\alpha(x) \in \mathcal{L}_1^\dagger$ with only one free variable, $\alpha(x)$ is equivalent to the translation of some formula $\varphi \in \mathcal{L}_s$ iff $\alpha(x)$ is invariant under l-bisimulation.

Proof The direction from left to right holds by Theorem 3.2 directly. We now begin to consider the other direction. Let $\alpha \in \mathcal{L}_1^\dagger$ with only one free variable. Suppose that α is invariant under l-bisimulation. Define $\mathbb{C}_l(\alpha) := \{\mathfrak{T}(\varphi, x, \varnothing, \varnothing) \mid \varphi \in \mathcal{L}_s \text{ and } \alpha \vDash \mathfrak{T}(\varphi, x, \varnothing, \varnothing)\}$. Each formula of $\mathbb{C}_l(\alpha)$ has only one free variable, i.e., x. We now show $\mathbb{C}_l(\alpha) \vDash \alpha$. Let $\langle \mathcal{M}, w \rangle \in \mathfrak{M}^\bullet$ such that $\mathcal{M} \vDash \mathbb{C}_l(\alpha)[\sigma_{x:=w}]$. First, we prove that $\Sigma = \mathfrak{T}((\mathbb{T}^l(\mathcal{M}, w), x, \varnothing, \varnothing)) \cup \{\alpha\}$ is consistent.

Suppose that Σ is not consistent. By the compactness of FOL, it holds that $\vDash \alpha \rightarrow \neg \bigwedge \Gamma$ for some finite $\Gamma \subseteq \mathfrak{T}(\mathbb{T}^l(\mathcal{M}, w), x, \varnothing, \varnothing)$. Then, from the definition of $\mathbb{C}_l(\alpha)$, we know $\neg \bigwedge \Gamma \in \mathbb{C}_l(\alpha)$, which is followed by $\neg \bigwedge \Gamma \in \mathfrak{T}(\mathbb{T}^l(\mathcal{M}, w), x, \varnothing, \varnothing)$. However, it contradicts to $\Gamma \subseteq \mathfrak{T}(\mathbb{T}^l(\mathcal{M}, w), x, \varnothing, \varnothing)$.

Now we show $\mathcal{M} \vDash \alpha[\sigma_{x:=w}]$. Since Σ is consistent, there is some $\langle \mathcal{M}', w' \rangle \in \mathfrak{M}^\bullet$ such that $\mathcal{M}' \vDash \Sigma[\sigma_{x:=w'}]$. Consequently, it holds that $\langle \mathcal{M}, w \rangle \leftrightsquigarrow_l \langle \mathcal{M}', w' \rangle$. Now take two ω-saturated elementary extensions $\langle \mathcal{M}_\omega, w \rangle$ and $\langle \mathcal{M}'_\omega, w' \rangle$ of $\langle \mathcal{M}, w \rangle$ and $\langle \mathcal{M}', w' \rangle$ respectively. By the invariance of FOL under elementary extensions, $\mathcal{M}' \vDash \alpha[\sigma_{x:=w'}]$ entails $\mathcal{M}'_\omega \vDash \alpha[\sigma_{x:=w'}]$. Moreover, by Theorem 3.3 and the assumption that α is invariant

for l-bisimulation, we have $\mathcal{M}_\omega \vDash \alpha[\sigma_{x:=w}]$. By the elementary extension, we obtain $\mathcal{M} \vDash \alpha[\sigma_{x:=w}]$. Therefore, $\mathbb{C}_l(\alpha) \vDash \alpha$.

Finally, we show that formula α is equivalent to the translation of an $\mathcal{L}_s$-formula. Since $\mathbb{C}_l(\alpha) \vDash \alpha$, from the compactness and deduction theorems of FOL, it follows that $\vDash \bigwedge \Gamma \to \alpha$ for some finite $\Gamma \subseteq \mathbb{C}_l(\alpha)$. Besides, by the definition of $\mathbb{C}_l(\alpha)$, we have $\vDash \alpha \to \bigwedge \Gamma$. Thus, $\vDash \alpha \leftrightarrow \bigwedge \Gamma$. Now the proof is completed. ∎

Therefore, in terms of the expressivity, CLL is as powerful as the one free variable fragment of FOL that is invariant for l-bisimulation.

3.4 Model checking and satisfiability for CLL

In this section, we consider the model checking problem and satisfiability problem for CLL. In particular, we show that the model checking problems for both CLL and $\mathcal{L}_{\diamondsuit\langle+\rangle}$ are PSPACE-complete. Also, both CLL and $\mathcal{L}_{\diamondsuit\langle-\rangle_{on}}$ lack the finite model property, and their satisfiability problems are undecidable.

Theorem 3.5: Model checking for logic CLL is PSPACE-complete.

Proof As mentioned, an upper bound can be established by Definition 3.5, which suggests that model checking for CLL is in PSPACE. On the other hand, a lower bound can be provided by a reduction f from BML into $\mathcal{L}_{\diamondsuit\langle+\rangle}$. Precisely, f is the reverse of the translation used in Proposition 3.3. Clearly, f has a polynomial size increase. Let $\langle W, R_L, V\rangle$ be a standard relational model and $w \in W$. It holds that $\langle W, R_L, V\rangle, w \vDash \varphi$ iff $\langle W, R_L, W \times W, V\rangle, w \vDash f(\varphi)$. Since the model checking problem for BML is also PSPACE-complete (Areces et al., 2015), the model checking for CLL is PSPACE-hard. The proof is completed. ∎

By the same reasoning as in the proof of Theorem 3.5, but now focusing on $\mathcal{L}_{\diamondsuit\langle+\rangle}$ instead of the whole CLL, we can obtain the following result:

Theorem 3.6: Model checking for $\mathcal{L}_{\diamondsuit\langle+\rangle}$ is PSPACE-complete.

Given the form of formula (3-1) describing winning strategies in correct learning games, it is also an interesting problem concerning the game to study the complexity of the model checking for the fragment of $\mathcal{L}_s$ consisting only of operators $\wedge, \boxdot, \langle-\rangle_{on}, \langle-\rangle_{on}$ and $\langle+\rangle$ (without $\neg$). We leave this as an open problem.

Now we move to considering the satisfiability problem. In particular, it will be shown that CLL is undecidable. To achieve this, in what follows we will study $\mathcal{L}_{\langle\cdot\rangle\langle-\rangle_{on}}$ instead of CLL. We first show that the fragment does not enjoy the finite model property. To prove this, our method is similar to that of Theorem 2.7, i.e., using the techniques of 'spy point', but details are very different in this new setting.

Theorem 3.7: $\mathcal{L}_{\langle\cdot\rangle\langle-\rangle_{on}}$ does not enjoy the finite model property.

Proof To prove this, we construct an $\mathcal{L}_{\langle\cdot\rangle\langle-\rangle_{on}}$-formula that can only be satisfied by some infinite models. Let φ_∞ be the conjunction of the following formulas:

$$
\begin{aligned}
(F_1)\quad & p \wedge q \wedge \langle\cdot\rangle p \wedge \langle\cdot\rangle \neg p \wedge [\cdot] \neg q \\
(F_2)\quad & [\cdot]\,(p \to \langle\cdot\rangle q \wedge \langle\cdot\rangle \neg q \wedge [\cdot] p) \\
(F_3)\quad & [\cdot]\,(p \to [\cdot](q \to [\cdot] \neg q \wedge \langle\cdot\rangle \neg p)) \\
(F_4)\quad & \langle\cdot\rangle(\neg p \wedge \langle-\rangle_{on} [\cdot]\,(p \wedge [\cdot](q \to [\cdot] p))) \\
(F_5)\quad & [\cdot]\,(p \to [\cdot](\neg q \to \langle\cdot\rangle q \wedge \langle\cdot\rangle \neg q \wedge [\cdot] p)) \\
(F_6)\quad & [\cdot]\,(p \to [\cdot](\neg q \to [\cdot](q \to [\cdot] \neg q \wedge \langle\cdot\rangle \neg p))) \\
(F_7)\quad & \langle\cdot\rangle(\neg p \wedge \langle-\rangle_{on} [\cdot]\, [\cdot](\neg q \to [\cdot](q \to [\cdot] p))) \\
(Spy)\quad & [\cdot]\,(p \to [\cdot](\neg q \to [\cdot](q \to \langle-\rangle_{on}(\neg q \wedge [\cdot] \neg q \wedge \langle-\rangle_{on}(q \wedge \langle\cdot\rangle(p \wedge [\cdot] \neg q)))))) \\
(Irr)\quad & [\cdot]\,(p \to [\cdot](q \to \langle-\rangle_{on}(\neg q \wedge [\cdot] \neg q \wedge [\cdot]\langle\cdot\rangle q))) \\
(No\text{-}3cyc)\quad & \neg\langle\cdot\rangle(p \wedge [\cdot](q \to \langle-\rangle_{on}(\neg q \wedge [\cdot](\neg q \wedge [\cdot](q \to \langle-\rangle_{on}(\neg q \wedge [\cdot] \neg q \wedge \\
& \langle-\rangle_{on}(q \wedge \langle\cdot\rangle(p \wedge [\cdot] \neg q)) \wedge \langle\cdot\rangle\langle\cdot\rangle(p \wedge [\cdot] \neg q))))))) \\
(Trans)\quad & [\cdot]\,(p \to [\cdot](q \to \langle-\rangle_{on}(\neg q \wedge [\cdot] \neg q \wedge [\cdot]\,[\cdot]\,(\neg q \to [\cdot](q \to \langle-\rangle_{on}(\neg q \wedge \\
& [\cdot]\,\neg q \wedge \langle-\rangle_{on}(p \wedge \neg\langle\cdot\rangle q \wedge \langle\cdot\rangle\,[\cdot]\,\neg q)))))))
\end{aligned}
$$

Formula φ_∞ is satisfiable (see Figure 3.4). Now we show that for any $\langle \mathcal{M}, w\rangle \in \mathfrak{M}^\bullet$, if $\mathcal{M}, w \vDash \varphi_\infty$, then $\mathcal{M}$ is infinite. Define $B := \{v \in W \mid v \in R_L(w) \cap V(p)\}$. In what follows, we assume that all previous conjuncts hold.

By (F_1), node w is $(p \wedge q)$, and $R_L(w) \cap V(q) = \varnothing$. Consequently, $\neg R_L ww$. Besides, $B \neq \varnothing$ and $R_L(w) \setminus B \neq \varnothing$. From (F_2), it follows that each element of B can see some $(q \wedge p)$-point(s) and $(\neg q \wedge p)$-point(s) via R_L, but cannot see any $\neg p$-points through R_L. Hence each point in B has at least one R_L-successor distinct from itself. By (F_3), for any $w_1 \in B$, each of its R_L-successors that is q can see some $\neg p$-point(s) via R_L, but cannot

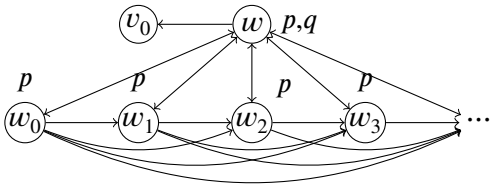


Figure 3.4 A model of formula φ_∞ (every link in the model belongs to R_L, and $R_T = \varnothing$). It can be shown that the formula is true at w.

see any q-points by R_L. By (F_4), $R_L(w) \setminus B \neq \varnothing$ is a singleton. Moreover, each $w_1 \in B$ can see point w via R_L, and for each $w_2 \in V(q)$, $R_L w_1 w_2$ entails $w_2 = w$.

Formulas (F_2)-(F_4) show the properties of the $(\neg q \wedge p)$-points accessible from w in one step by R_L. Similarly, formulas (F_5)-(F_7) play the same roles as (F_2)-(F_4) respectively, but focusing on showing the properties of the $(\neg q \wedge p)$-points accessible from w in 2 steps via R_L. In particular, (F_7) guarantees that every $(\neg q \wedge p)$-point w_1 accessible from w in 2 steps by R_L can also see w via R_L, and that for each q-point w_2, $R_L w_1 w_2$ entails $w_2 = w$.

Formula (*Spy*) shows that, for any $(\neg q \wedge p)$-points w_1, w_2 such that $R_L w w_1$ and $R_L w_1 w_2$, after removing some $\langle v, v' \rangle \in \{\langle w, w_1 \rangle, \langle w_1, w_2 \rangle, \langle w_2, w \rangle\}$, state v is $\neg q$ and does not have any q-successors. As $w \in V(q)$, we have $v \neq w$. Besides, if $\langle v, v' \rangle = \langle w_1, w_2 \rangle$, after we cut $\langle v, v' \rangle$, v still can see $w \in V(q)$, so $\langle v, v' \rangle = \langle w_2, w \rangle$. Also, after deleting $\langle w, w_1 \rangle$, point w can reach a p-point w_3 via R_L such that $R_L(w_3) \cap V(q) = \varnothing$. Therefore, $w_3 = w_2$. Thus, (*Spy*) ensures that each $(\neg q \wedge p)$-point w_1 accessible from w in 2 steps via R_L is also accessible from w in one step via R_L.

By (*Irr*), for each $w_1 \in B$, it holds $\neg R_L w_1 w_1$. Formula (*No-3cyc*) shows R_L-links cannot be cycles of length 2 or 3 in B, and (*Trans*) forces R_L to transitively order B.

Hence $\langle B, R_L \rangle$ is an unbounded strict partial order, thus B is infinite and so is W. This completes the proof. ∎

We now proceed to show the undecidability of $\mathcal{L}_{\diamondsuit\langle - \rangle_{on}}$, by encoding the $\mathbb{N} \times \mathbb{N}$ tiling problem. In what follows, we will use three modalities $\diamondsuit_s$, $\diamondsuit_u$ and $\diamondsuit_r$ to stand for $\diamondsuit$. Correspondingly, a model $\mathcal{M} = \{W, R_s, R_u, R_r, R_T, V\}$ now includes four relations. We are going to construct a spy point over R_s, and relations R_u, R_r represent moving up and to the right, respectively, from one tile to the other. Intuitively, the union of these three relations can be treated as R_L in the model. Moreover, as illustrated by the following proof, they are disjoint from each other. So they are a partition of R_L. Thanks to this, we do not need any extra modalities to represent $\langle - \rangle_{on}$.

Theorem 3.8: The satisfiability problem for $\mathcal{L}_{\diamondsuit\langle-\rangle_{on}}$ is undecidable.[1]

Proof Let $T = \{T_1, \cdots, T_n\}$ be a finite set of tile types. Again, for each T_i, we denote by $u(T_i)$, $d(T_i)$, $l(T_i)$ and $r(T_i)$ respectively the colors of its up, down, left and right edges. Also, each T_i is coded with a fixed proposition t_i. Now we show that φ_T, the conjunction of the following formulas, is true iff T tiles $\mathbb{N} \times \mathbb{N}$.

$$(M_1) \quad p \wedge q \wedge \diamondsuit_s p \wedge \diamondsuit_s \neg p \wedge \boxdot_s \neg q \wedge \diamondsuit_s \langle - \rangle_{on} \boxdot_s p$$

$$(M_2) \quad \boxdot_s (p \rightarrow \diamondsuit_s \top \wedge \boxdot_s (q \wedge \diamondsuit_s \neg p))$$

$$(M_3) \quad \diamondsuit_s (\neg p \wedge \langle - \rangle_{on} \boxdot_s \boxdot_s (q \wedge \neg \diamondsuit_s \neg p))$$

$$(M_4) \quad \boxdot_s (p \rightarrow \diamondsuit_u \top \wedge \boxdot_u (p \wedge \neg q \wedge \diamondsuit_s \top \wedge \boxdot_s (q \wedge \diamondsuit_s \neg p)))$$

$$\boxdot_s (p \rightarrow \diamondsuit_r \top \wedge \boxdot_r (p \wedge \neg q \wedge \diamondsuit_s \top \wedge \boxdot_s (q \wedge \diamondsuit_s \neg p)))$$

$$(M_5) \quad \diamondsuit_s (\neg p \wedge \langle - \rangle_{on} \boxdot_s \boxdot_u \boxdot_s \neg \diamondsuit_s \neg p)$$

$$\diamondsuit_s (\neg p \wedge \langle - \rangle_{on} \boxdot_s \boxdot_r \boxdot_s \neg \diamondsuit_s \neg p)$$

$$(M_6) \quad \boxdot_s (p \rightarrow \boxdot_u (\diamondsuit_u \top \wedge \diamondsuit_r \top \wedge \boxdot_u (p \wedge \neg q) \wedge \boxdot_r (p \wedge \neg q)))$$

$$\boxdot_s (p \rightarrow \boxdot_r (\diamondsuit_u \top \wedge \diamondsuit_r \top \wedge \boxdot_u (p \wedge \neg q) \wedge \boxdot_r (p \wedge \neg q)))$$

$$(M_7) \quad \boxdot_s (p \rightarrow \boxdot_s (q \wedge \langle - \rangle_{on} (\neg q \wedge \boxdot_u (\diamondsuit_s q \wedge \neg \diamondsuit_u \neg \diamondsuit_s q))))$$

$$\boxdot_s (p \rightarrow \boxdot_s (q \wedge \langle - \rangle_{on} (\neg q \wedge \boxdot_r (\diamondsuit_s q \wedge \neg \diamondsuit_r \neg \diamondsuit_s q))))$$

$$(Spy) \quad \boxdot_s (p \rightarrow \boxdot_u \boxdot_s \langle - \rangle_{on} (\boxdot_s \bot \wedge \langle - \rangle_{on} (p \wedge q \wedge \diamondsuit_s (p \wedge \boxdot_s \bot))))$$

$$\boxdot_s (p \rightarrow \boxdot_r \boxdot_s \langle - \rangle_{on} (\boxdot_s \bot \wedge \langle - \rangle_{on} (p \wedge q \wedge \diamondsuit_s (p \wedge \boxdot_s \bot))))$$

$$(Func) \quad \boxdot_s (p \rightarrow \boxdot_s \langle - \rangle_{on} (\boxdot_s \bot \wedge \boxdot_u \langle - \rangle_{on} (\boxdot_s \bot \wedge \boxdot_u \bot))$$

$$\boxdot_s (p \rightarrow \boxdot_s \langle - \rangle_{on} (\boxdot_s \bot \wedge \boxdot_r \langle - \rangle_{on} (\boxdot_s \bot \wedge \boxdot_r \bot))$$

$$(No\text{-}UR) \quad \boxdot_s (p \rightarrow \boxdot_s \langle - \rangle_{on} (\boxdot_s \bot \wedge \boxdot_u \boxdot_r \diamondsuit_s q \wedge \boxdot_r \boxdot_u \diamondsuit_s q))$$

$$(No\text{-}URU) \quad \boxdot_s (p \rightarrow \boxdot_s \langle - \rangle_{on} (\boxdot_s \bot \wedge \boxdot_u \boxdot_r \boxdot_u \diamondsuit_s q))$$

$$(Conv) \quad \boxdot_s (p \rightarrow \boxdot_s \langle - \rangle_{on} (\boxdot_s \bot \wedge \diamondsuit_u \boxdot_s \langle - \rangle_{on} (\boxdot_s \bot \wedge \diamondsuit_u \top \wedge$$

$$\diamondsuit_r \boxdot_u \langle - \rangle_{on} (\boxdot_u \bot \wedge \diamondsuit_s \diamondsuit_s (p \wedge \boxdot_s \bot \wedge \diamondsuit_r \diamondsuit_u (p \wedge \boxdot_u \bot))))))$$

$$(Unique) \quad \boxdot_s (p \rightarrow \bigvee_{1 \leqslant i \leqslant n} t_i \wedge \bigwedge_{1 \leqslant i < j \leqslant n} (t_i \rightarrow \neg t_j))$$

$$(Vert) \quad \boxdot_s (p \rightarrow \bigwedge_{1 \leqslant i \leqslant n} (t_i \rightarrow \diamondsuit_u \bigvee_{1 \leqslant j \leqslant n,\, u(T_i) = d(T_j)} t_j))$$

1 Similar to the case of Theorem 2.8, the four modalities used in its proof can be reduced to two by a standard argument, but we will omit the details because of the syntactic cost involved in writing the formulas.

$$(\mathit{Horiz}) \qquad \boxdot_s (p \to \bigwedge_{1 \leqslant i \leqslant n} (t_i \to \diamondsuit_r \bigvee_{1 \leqslant j \leqslant n,\ r(T_i) = l(T_j)} t_j))$$

Let $\mathcal{M} = \{W, R_s, R_u, R_r, R_T, V\}$ be a model and $w \in W$ such that $\mathcal{M}, w \vDash \varphi_T$. We show that $\mathcal{M}$ tiles $\mathbb{N} \times \mathbb{N}$. Define $G := \{v \in W \mid v \in R_s(w) \cap V(p)\}$ where $R_s(w) := \{v \in W \mid R_s w v\}$. We will use the elements of G to represent tiles.

By formula (M_1), node w is $(p \wedge q)$, and $R_s(w) \cap V(q) = \varnothing$. So, it holds that $\neg R_s w w$. Besides, $R_s(w) \setminus G$ is a singleton (e.g., $\{v\}$) and $G \neq \varnothing$. By (M_2), each tile w_1 has some successor(s) via R_s, and each $w_2 \in R_s(w_1)$ is q and has some $\neg p$-successor(s) via R_s. Formulas (M_1) and (M_2) illustrate that R_s is irreflexive. Formula (M_3) ensures that each tile w_1 can see w via R_s and that for each $w_2 \in V(q)$, $R_s w_1 w_2$ entails $w_2 = w$. From (M_4), we know that each tile has some successor(s) via R_u and some successor(s) via R_r. Besides, each point accessible from a tile via R_u or R_r is $(\neg q \wedge p)$, and it has some q-successor(s) w_1 via R_s where each w_1 can see some $\neg p$-point(s) via R_s. By formula (M_5), each $w_1 \in W$ accessible from a tile via R_u or R_r can see w by R_s. Also, for each $(q \wedge p)$-point w_2, if $w_2 \in R_s(w_1)$, then $w_2 = w$. Formula (M_6) ensures that each $w_1 \in W$ accessible from some tile via R_u or R_r also has some successor(s) via R_u and some successor(s) via R_r. Besides, each such successor via R_u or R_r is $(\neg q \wedge p)$. Formula (M_7) shows that both the restrictions of R_u and R_r to $G \times G$ are irreflexive and asymmetric. By (*Spy*), w is a spy point via R_s. Note that formula (M_4) says that each tile has some tile(s) above it and some tile(s) to its right. Now, from formula (*Func*), we know that each tile has exactly one tile above it and exactly one tile to its right. By (*No-UR*), any tile cannot be above/below as well as to the left/right of another tile. Formula (*No-URU*) disallows cycles following successive steps of the R_u, R_r, and R_u relations, in this order. Moreover, (*Conv*) ensures that the tiles are arranged as a grid. Formula (*Unique*) guarantees that each tile has a unique type. Finally, (*Vert*) and (*Horiz*) force the colors of tiles to match properly. Thus, $\mathcal{M}$ tiles $\mathbb{N} \times \mathbb{N}$.

On the other hand, it is easy to see that any tiling of $\mathbb{N} \times \mathbb{N}$ induces a model for φ_T. Now the proof is completed. ∎

From Theorem 3.7 and Theorem 3.8, it follows immediately that logic CLL lacks the finite model property, and its satisfiability problem is undecidable.

Finally, it is worth noting that, besides $\mathcal{L}_{\diamondsuit\langle - \rangle_{on}}$, other fragments also deserve to be studied, say, $\mathcal{L}_{\diamondsuit\langle - \rangle_{off}}$. It is already known that the satisfiable problem for SML is undecidable (Aucher et al., 2018) and its model checking problem is PSPACE-complete (Areces

et al., 2015). Given the similarity between $\langle - \rangle_{\mathit{off}}$ and $\langle - \rangle$ (recall Proposition 3.2), is the model checking for $\mathcal{L}_{\Diamond\langle - \rangle_{\mathit{off}}}$ PSPACE-complete? And is its satisfiability problem undecidable?

3.5 Summary and future work

Summary. Using graph games, the chapter investigated the interactions of agents in teaching/learning scenarios. Of course, there is much more to learning and teaching than simple link deletion or addition. But our simple proposal of graph games was still powerful enough to highlight a number of realistic features of learning/teaching processes. Compared with sabotage games, the analysis presented in this chapter enabled us to capture the following further aspects:

- In the process of learning, Learner may make mistakes and ignore the correct relation between different hypotheses, while Teacher could help her to correct mistakes and point out facts ignored.
- We distinguished those potential mistakes from actual ones, which in turn helped us to characterize the subtle differences between their eliminations.
- We placed the success condition of learning on a solid foundation: reaching the right conclusion cannot be by chance, and the process itself should also be coherent. In this way, our framework excluded Gettier cases from the success in learning.

After presenting the graph games, we also determined a logical system that was able to characterize precisely players' actions and winning positions. Moreover, we also explored many meta-properties of the logic, such as

- We provided some interesting observations and logical validities, which were useful for us to understand some basics of the device.
- We studied basics of its expressivity, including its first-order translation, a novel notion of bisimulation and a characterization theorem for CLL as a fragment of FOL that is invariant under the bisimulation introduced.
- It was shown that model checking for CLL is PSPACE-complete, the logic did not enjoy the finite model property, and its satisfiability problem was undecidable.

Relevant research. This work takes a small step towards studying the interaction between graph games, logics and formal learning theory, a major tool in formal epistemology. As mentioned, the success condition of learning used in this chapter is finite identification.

Both Mukouchi (1992) and Lange et al. (1996) study this kind of learnability in the context of indexed families of recursive languages. More generally, a relaxed (and more common) notion of finite identification in the limit and its relation to logics of information update and belief change is studied in (Dégremont and Gierasimczuk, 2011; Gierasimczuk, 2009a,b; Gierasimczuk and de Jongh, 2013).

This chapter is also closely related to the existing work on graph games and logics. We have already gave detailed discussion on this in Section 2.7. Moreover, another relevant and congenial line of research is epistemic logics. As mentioned already, one goal of our work in this chapter has been to offer a perhaps new game-theoretic angle on avoiding the notorious Gettier problem in analyzing knowledge.[1]

Future work. This chapter just made a start, and many problems remain to be studied at the interface of logic, games and learning theory.

Starting with logical issues, Section 3.2.2 showed that the system CLL can express winning positions for players in finite learning/teaching games, but to capture those appropriate to infinite games of never-ending learning, can CLL be expanded, say, with fixpoint operators? From the translation in Definition 5, we know that CLL is effectively axiomatizable (cf. van Benthem, 1984). However, is it possible to axiomatize the logic presented above as a base theory of learning and teaching in a perspicuous Hilbert-style calculus? Finally, Proposition 5 shows that the validities of CLL are not closed under substitution. But are the schematic validities of CLL axiomatizable, perhaps even decidable?

In terms of games, we do not know the complexity of CLG, although we gave a basic observation on the necessary conditions for winning. Besides, CLG includes exactly two players, and it is meaningful to study (classroom-style) settings with more agents. Also, CLG is a cooperative game, but there are also other scenarios closer to actual abilities and attitudes of players. Say, Learner may be unwilling to learn, and Teacher can also be unhelpful or not omniscient. What would be natural variants of CLG capturing these features of actual teaching and learning?[2]

1 For another approach to this famous challenge, compare Baltag et al. (2019a) on a use of topological semantics to analyze the notion of *full belief*.

2 Many further structures deserve investigation. E.g., consider the significance of *cycles* in CLG. Suppose that Learner reaches the goal through a path $\langle a_0, a_1, ..., a_i, ..., a_m, ..., a_n\rangle$ from the starting node a_0 to the goal node a_n, where $Set(\langle a_0, a_1, ..., a_i\rangle) \cup Set(\langle a_m ..., a_n\rangle) \subseteq R_T$, $Set(\langle a_i ..., a_m\rangle) \nsubseteq R_T$ and $a_i = a_m$. According to our theory, Learner has not reached the goal through a correct path, so she has not learnt properly. However, it can be argued that even if one learns some redundant circular argument in addition to a proper argument, one has still learned a good deal. But our game now cannot capture these scenarios. A possible solution is to define learning such that it could include 'meaningless' cycles.

Finally, although Section 3.1 discusses some applications of CLG to learning, much more needs to be done. As we have noted at several places already, the relations between, or fruitful combinations of, our game-logical framework and standard formal learning theory deserve to be studied more systematically.[1]

Even despite all these further issues emerging from our work and the many desiderata yet to be fulfilled, we hope to have shown in this chapter that a game logic perspective on learning and teaching is both feasible and attractive.

1 For more discussions on the applications of SG-style frameworks to paradigms of learning theory, we refer to Gierasimczuk et al. (2009), whose arguments also apply to CLG after minor modifications.

Chapter 4 Logical proposals for dynamic dependence

4.1 Motivation: dynamic dependence in graph games

The preceding two chapters showed how graph games provide a flexible approach to different sorts of social interactions in various contexts. These games also had natural matching logics that encode reasoning about players' goals and strategic abilities. Even so, reflecting the particular patterns of interaction that were studied, our logical frameworks had surprisingly high complexity: their theories of modifying models turned out to be undecidable. We now move away from such detailed tools for analysis to look at general features of game-like social situations. Exploring fundamental aspects of interaction in this abstract way may provide another form of insight into social interactions, which may eventually also inspire new types of modeling with lower complexity.

In this chapter, we study one significant general feature of social interaction: *dynamic dependence*, i.e., dependence between actions of social agents manifesting itself over time. From the perspective of our earlier games, strategic dependence arises as follows:

Dependence between actions: How a player acts now depends on what the opponent has or has not done previously (cf. e.g., Example 1.3).

Before introducing details of our work, it is important to be pointed out that although the chapter is motivated by the interactions of agents in game-like scenarios, dependence itself is a very basic notion in many areas, such as counterfactual reasoning, database theory, dynamical systems theory, game theory, or social science. For instance, an agent in a social network may adopt a given behavior or opinion in the next step depending on the proportion of friends who currently have that behavior, and a global evolution of behaviors for the whole group then unfolds in a dynamical system (cf. e.g., Baltag et al., 2019b). So far, logicians have mostly studied static dependencies (as well as matching forms of static independence) in various frameworks, (cf. e.g., Hintikka, 1998; Väänänen, 2007). We now proceed to add temporal aspects.

Our analysis will focus on dynamical systems whose states are assignments of values to variables, as in the semantics of first-order logic. Crucially, not all such functions need occur as states as the system evolves over time, and this is what leads to dependencies: a change in value for one variable may only be possible by also changing the value of some

other variable.[1] This correlating feature of 'assignment gaps' is well-known from logical analyses of dependence (Andréka et al., 1998; Hodges, 1997).

To talk about dynamic dependencies that manifest themselves over time, we proceed in two stages. Our first analysis combines two systems, (i) the modal logic LFD of instantaneous functional dependence from Baltag and van Benthem (2021b), and (ii) basic vocabulary from temporal logic. Here LFD gives us dependence atoms $D_X y$ expressing dependence of the current value of y on the current values of the variables in X, plus modalities $\mathsf{D}_X\varphi$ that express which facts are forced to be true by the current values of the variables in X. From temporal logic, we take the operator $\bigcirc\varphi$ stating that φ is true at the next state produced by the transition function of the dynamical system. Connecting the two components, we will use 'dynamic dependence formulas' $D^n_X y$ saying that the value of y in n steps from now depends functionally on the current values of the variables X.

The resulting logical system DFD can express interesting facts about temporal dependence, and as we shall show, it has a complete axiomatization. Our methods for proving this resemble known ones for LFD and temporal logics, but the combination is surprisingly challenging from a technical perspective, as will become clear in what follows. Yet we feel that at least this much effort is needed to make good on the suggestion in Baltag and van Benthem (2021b) that the analysis for the timeless case might carry over to causality and games.

Next, dynamical systems usually come with a topology on the state space that plays an essential role in evolution over time. Accordingly, the transition function driving the system is often continuous w.r.t. this topology, and we must deal with forms of *continuous dependence*. To access this further structure, we use a richer temporal-topological logic of dynamical systems in the style of (Artemov et al., 1997; Kremer and Mints, 2007). In particular, this language has a basic topological modality $\Box\varphi$ saying that φ is true throughout some open neighborhood of the current state. The resulting system DCD can be seen as a natural extension of dynamic topological logics with dependence structure, a topic of interest in its own right. This time, in addition to the technical challenges occurring with DFD, there are also conceptual challenges in finding the right notions. Continuous dependence can mean various things, and though we settle on one particular proposal that we consider general and useful, and for which we can prove results, we have only made a

1 From the perspective of our games, this intuitively can be taken as the effect of the rationality of players which would stop them to take some 'bad' actions: when a player behaves in another way, the other player will also change her actions.

start.

The chapter is organized as follows. Section 4.2 lays out the basics of DFD, including two equivalent but complementary versions of its semantics. Section 4.3 presents a complete Hilbert-style proof system for the logic and shows the decidability of a significant fragment of our logic. Next, Section 4.4 moves towards a topological setting and analyzes the basic logic DCD of dynamic continuous dependence. Section 4.6 concludes and points at directions for further research.

4.2 The logic DFD: language and semantics

4.2.1 Language

Let us first fix the language for this chapter. Take a finite set of variables $\mathbb{V}$ and a vocabulary of predicate symbols $Pred$ with arities given by $ar : Pred \to \mathbb{N}$.[1]

Definition 4.1: The *language* $\mathcal{L}_{\mathsf{D}}$ is given by the grammar

$$\varphi ::= P\mathbf{x} \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \bigcirc\varphi \mid \mathsf{D}_X\varphi \mid D^n_X y$$

where $n \in \mathbb{N}$ is a natural number, $y \in \mathbb{V}$ is a variable, P is a predicate symbol, $\mathbf{x} = \langle x_1, \cdots, x_n\rangle$ is a sequence of variables of length $n = ar(P)$, and $X \subseteq \mathbb{V}$. Moreover, $\bot$, $\top$, $\vee$, $\to$ and $\leftrightarrow$ are defined as usual, while $D^n_X Y := \bigwedge_{y\in Y} D^n_X y$, and $\bigcirc^n\varphi$ is the n-th iteration of $\bigcirc$. Finally, we write $D_X y$ for $D^0_X y$, $D_x y$ for $D_{\{x\}} y$, $\mathsf{D}_x\varphi$ for $\mathsf{D}_{\{x\}}\varphi$, and $\widehat{\mathsf{D}}_X\varphi$ for $\neg\mathsf{D}_X\neg\varphi$.

This is like an enriched first-order language, now in a temporal setting where atomic formulas $P\mathbf{x}$ say that the current values of $\mathbf{x}$ satisfy the predicate P.[2]

The semantics for $\mathcal{L}_{\mathsf{D}}$ comes in two flavors: in Section 4.2.2, we present a format of generalized first-order models, which allows us to exploit a technical translation into FOL presented in Section 4.2.3. However, in Section 4.2.4, we reformulate the semantics in terms of relational models for modal logic, which allows us to also use techniques from modal logic. As shown in Section 4.2.5, the two types of models are semantically equivalent, but it is useful to think in both styles.

1 Finite sets of variables are common in practice. Dealing with an infinite set of variables leads to some complexities in our system that we decided to leave aside here.

2 In line with notations '$x(t)$, $x(t+1)$' in dynamical systems, one might also let $\bigcirc$ operate on variables to denote their future values, and write dependence atoms $D^n_X y$ as $D_{\bigcirc^n} y$. This notation would give a slightly different guise for the logic to follow.

4.2.2 First-order semantics

Definition 4.2: A *dynamic dependence model* is a tuple $\mathfrak{M} = \langle M, A, g\rangle$, where

- $M = \langle O, I\rangle$ is a model of FOL with object domain O and interpretation map I assigning sets $I(P)$ of n-tuples of objects to n-ary predicate symbols P.
- A is a set of admissible assignments.
- $g : A \to A$ is a total function.

Here, unlike with standard FOL, not all assignments are available: we only consider those in A. The function g is the next-state map for the dynamical system, with finite iterations written as $g^n(s)$. Finally, for $X \subseteq \mathbb{V}$ and $s, t \in A$, we write $s =_X t$ when $s(x) = t(x)$ for all $x \in X$. Here is our semantics.

Definition 4.3: Given a dynamic dependence model $\mathfrak{M} = \langle M, A, g\rangle$, *truth of a formula* $\varphi \in \mathcal{L}_{\mathsf{D}}$ *in* $\mathfrak{M}$ *at* $s \in A$, written $\mathfrak{M}, s \vDash \varphi$, is defined as follows:

$$
\begin{array}{rll}
\mathfrak{M}, s \vDash Px_1 \cdots x_n & \text{iff} & \langle s(x_1), s(x_2), \cdots, s(x_n)\rangle \in I(P) \\
\mathfrak{M}, s \vDash \neg\varphi & \text{iff} & \text{not } \mathfrak{M}, s \vDash \varphi \\
\mathfrak{M}, s \vDash (\varphi \wedge \psi) & \text{iff} & \mathfrak{M}, s \vDash \varphi \text{ and } \mathfrak{M}, s \vDash \psi \\
\mathfrak{M}, s \vDash \bigcirc\varphi & \text{iff} & \mathfrak{M}, g(s) \vDash \varphi \\
\mathfrak{M}, s \vDash D^n_X y & \text{iff} & \text{for each } t \in A,\ s =_X t \text{ implies } g^n(s) =_y g^n(t) \\
\mathfrak{M}, s \vDash \mathsf{D}_X\varphi & \text{iff} & \text{for each } t \in A,\ s =_X t \text{ implies } \mathfrak{M}, t \vDash \varphi
\end{array}
$$

The model $\mathfrak{M}$ is often omitted when it is clear from context. For brevity in formulations, we will often write $(P\mathbf{x})^s$ for $s \vDash P\mathbf{x}$, and $(D^n_X y)^s$ for $s \vDash D^n_X y$.

The special case $\mathsf{D}_\varnothing\varphi$ is of importance in its own right, since it expresses the *universal modality* saying that φ is true at every assignment in the model.

Our dependence quantifiers $\mathsf{D}_X\varphi$ work differently from FOL quantifiers $\forall, \exists$: they 'free' variables rather than bind them. This feature better fits basic reasoning about dependencies, and the modalities $\mathsf{D}_X\varphi$ can define the first-order quantifiers given that $\mathfrak{V}$ is finite, Baltag and van Benthem (2021b). In this sense, the logic DFD generalizes standard first-order logic and the models over which it is interpreted.

There is in fact a special kind of variables in formulas of DFD that play a central role similar to that of the free variables in FOL.

Definition 4.4: Let φ be an $\mathcal{L}_{\mathsf{D}}$-formula. The set of *free variables* occurring in φ, $Free(\varphi)$, is defined recursively as follows:

- $Free(Px_1 \cdots x_n) = \{x_1, \cdots, x_n\}$
- $Free(\neg\varphi) = Free(\bigcirc\varphi) = Free(\varphi)$
- $Free(\varphi \wedge \psi) = Free(\varphi) \cup Free(\psi)$
- $Free(D^n_X y) = Free(\mathsf{D}_X\varphi) = X$

The key property of this new notion is the following *Locality Lemma*, which can be proved by formula induction using the recursive definition of $Free(\varphi)$:

Proposition 4.1: Let φ be an $\mathcal{L}_{\mathsf{D}}$-formula with nesting depth at most n for the temporal operator $\bigcirc$ and $Free(\varphi) \subseteq X$ for some set of variables X. For any dynamic dependence model $\mathfrak{M}$ and any two admissible assignments s, t with $g^m(s) =_X g^m(t)$ for all m with $0 \leqslant m \leqslant n$, it holds that $\mathfrak{M}, s \vDash \varphi$ iff $\mathfrak{M}, t \vDash \varphi$.

There is also a more technical connection of our logic with first-order logic. One can translate the language of DFD over dynamic dependence models into FOL over its standard models, in the style of Andréka et al. (1998); Baltag and van Benthem (2021b). This translation establishes, e.g., compactness and recursive axiomatizability for DFD via these properties for FOL. The translation needs care: details are in the following subsection.

4.2.3 First-order translation for DFD

An effective first-order translation for the logic DFD needs some syntactic ingredients to encode the relevant structure of dynamic dependence models $\mathfrak{M} = \langle M, A, g\rangle$. We enumerate the set $\mathbb{V}$ of variables as $\mathbf{v} = \langle x_1, \cdots, x_k\rangle$. Now, take fresh copies $\{x'_1, \cdots, x'_k\}$ and let $\mathbf{v}'$ the corresponding enumeration, i.e., y is the n-th variable of $\mathbf{v}$ iff y' is the n-th variable in $\mathbf{v}'$. Also, extend the language with a new k-ary predicate symbol A, where $A\mathbf{v}$ states intuitively that the tuple of values assigned to $\mathbf{v}$ belongs to the admissible assignments of the relevant dynamic dependence model. Moreover, to encode the dynamic transitions between assignments, we add k new k-ary functions g_i on variables. For a k-tuple $\mathbf{v}^*$ of variables, $g_i(\mathbf{v}^*)$ represents the value of the i-th element of $\mathbf{v}^*$ at the next step. For brevity, we write $g(\mathbf{v}^*)$ for the tuple $\langle g_1(\mathbf{v}^*), \cdots, g_k(\mathbf{v}^*)\rangle$, $g^n(\mathbf{v}^*)$ for the n-th iteration of g, and $g^n_i(\mathbf{v}^*)$ for the i-th element of $g^n(\mathbf{v}^*)$.

Finally, we denote by $\mathcal{T}(\mathfrak{M})$ the FOL model for the new expanded language that is naturally associated with a given dependence model $\mathfrak{M}$.

Definition 4.5: Let $\mathbf{v}^*$ a tuple obtained by replacing zero or more variables x_i in $\mathbf{v}$ with $g_i^j(\mathbf{v})$ or $g_i^j(\mathbf{v}')$. The *first-order translation* $\langle \mathcal{T}, g^m \rangle$ from $\mathcal{L}_{\mathsf{D}}$ to first-order formulas in the finite vocabulary introduced above is defined as follows:

(1) $\langle \mathcal{T}, g^m(\mathbf{v}^*)\rangle(Px_{i_1} \cdots x_{i_j}) = \forall \mathbf{z}(A\mathbf{v}^\dagger \rightarrow Pg_{i_1}^m(\mathbf{v}^*) \cdots g_{i_j}^m(\mathbf{v}^*)$, where $\mathbf{z}$ is the enumeration of variables in $\mathbb{V} \setminus \{x_{i_1}, \cdots, x_{i_j}\}$ and $\mathbf{v}^\dagger$ is the result of replacing each $x_{i_n} \in \{x_{i_1}, \cdots, x_{i_j}\}$ with $g_{i_n}^m(\mathbf{v}^*)$ in the enumeration $\mathbf{v}$.

(2) $\langle \mathcal{T}, g^m(\mathbf{v}^*)\rangle(\neg\varphi) = \neg\langle \mathcal{T}, g^m(\mathbf{v}^*)\rangle(\varphi)$

(3) $\langle \mathcal{T}, g^m(\mathbf{v}^*)\rangle(\varphi \wedge \psi) = \langle \mathcal{T}, g^m(\mathbf{v}^*)\rangle(\varphi) \wedge \langle \mathcal{T}, g^m(\mathbf{v}^*)\rangle(\psi)$

(4) $\langle \mathcal{T}, g^m(\mathbf{v}^*)\rangle(\bigcirc\varphi) = \langle \mathcal{T}, g^{m+1}(\mathbf{v}^*)\rangle(\varphi)$

(5) $\langle \mathcal{T}, g^m(\mathbf{v}^*)\rangle(\mathsf{D}_X\varphi) = \forall \mathbf{z}(A\mathbf{v}^+ \rightarrow \langle \mathcal{T}, g^0(\mathbf{v}^+)\rangle(\varphi))$, where $\mathbf{z}$ is the enumeration of variables in $\mathbb{V} \setminus X$ and $\mathbf{v}^+$ is the result of replacing each $x_i \in X$ with $g_i^m(\mathbf{v}^*)$ in the enumeration $\mathbf{v}$.

(6) $\langle \mathcal{T}, g^m(\mathbf{v}^*)\rangle(D_X^n x_i) = \forall \mathbf{z}\forall \mathbf{z}'(A\mathbf{v}^+ \wedge A\mathbf{v}^+[\mathbf{z}'/\mathbf{z}] \rightarrow g_i^n(\mathbf{v}^+) = g_i^n(\mathbf{v}^+[\mathbf{z}'/\mathbf{z}]))$, where $\mathbf{v}^+$ and $\mathbf{z}$ are as in (5), and $\mathbf{z}'$ is the corresponding $\mathbb{V}'$-copies of $\mathbf{z}$.

For each $\varphi \in \mathcal{L}_{\mathsf{D}}$ over the finite set $\mathbb{V}$, the free variables in the usual first-order sense in its translation $\langle \mathcal{T}, g^0(\mathbf{v})\rangle(\varphi)$ are exactly the above set $Free(\varphi)$.[1] Also, we can prove the correctness of the translation, in the sense of the following:

Proposition 4.2: For all dynamic dependence models $\mathfrak{M}$ and $\mathcal{L}_{\mathsf{D}}$-formulas φ:

$$\mathfrak{M}, s \vDash \varphi \text{ iff } \mathcal{T}(\mathfrak{M}), s \vDash \langle \mathcal{T}, g^0(\mathbf{v})\rangle(\varphi).$$

Proof We use induction on formulas φ. Here are the non-routine cases.

(1). $\varphi = P\mathbf{x}$. Then, by Definition 4.5, $\langle \mathcal{T}, g^0(\mathbf{v})\rangle(\varphi) = \forall \mathbf{z}(A\mathbf{v} \rightarrow P\mathbf{x})$, where $\mathbf{z}$ enumerates all variables not in $\mathbf{x}$. Now, by the truth definition, $\mathfrak{M}, s \vDash P\mathbf{x}$ iff all admissible assignments assigning the same values as s to variables in $\mathbf{x}$ make $P\mathbf{x}$ true. Thus, $\mathfrak{M}, s \vDash \varphi$ iff $\mathcal{T}(\mathfrak{M}), s \vDash \langle \mathcal{T}, g^0(\mathbf{v})\rangle(\varphi)$.

(2). $\varphi = \bigcirc\psi$. Let t be an admissible assignment s.t. $t = g(s)$. Then we have $\mathfrak{M}, s \vDash \bigcirc\psi$ iff $\mathfrak{M}, t \vDash \psi$. By the inductive hypothesis, $\mathfrak{M}, t \vDash \psi$ iff $\mathcal{T}(\mathfrak{M}), t \vDash \langle \mathcal{T}, g^0(\mathbf{v})\rangle(\psi)$. Also, the values of variables in $g^0(\mathbf{v})$ at t are those for the tuple $g^1(\mathbf{v})$ at s. Hence, $\mathcal{T}(\mathfrak{M}), t \vDash \langle \mathcal{T}, g^0(\mathbf{v})\rangle(\psi)$ iff $\mathcal{T}(\mathfrak{M}), s \vDash \langle \mathcal{T}, g^1(\mathbf{v})\rangle(\psi)$.

(3). $\varphi = \mathsf{D}_X\psi$. By our semantics, $\mathfrak{M}, s \vDash \mathsf{D}_X\psi$ iff for each admissible assignment t assigning the same values to X as s, $\mathfrak{M}, t \vDash \psi$. By the inductive hypothesis, the latter

1 Readers may have wondered why we did not define $\langle \mathcal{T}, g^m(\mathbf{v}^*)\rangle(Px_{i_1} \cdots x_{i_j})$ directly as $Pg_{i_1}^m(\mathbf{v}^*) \cdots g_{i_j}^m(\mathbf{v}^*)$. However, if we define the translation in this way, the set of free variables in $\langle \mathcal{T}, g^0(\mathbf{v})\rangle(P\mathbf{x})$ will be the whole set $\mathfrak{V}$ rather than those in $\mathbf{x}$.

says that $\mathcal{T}(\mathfrak{M}), s \vDash \forall \mathbf{z}(A\mathbf{v} \to \langle \mathcal{T}, g^0(\mathbf{v})\rangle(\psi))$, where $\mathbf{z}$ enumerates the variables in $\mathbb{V} \setminus X$. As formula $\forall \mathbf{z}(A\mathbf{v} \to \langle \mathcal{T}, g^0(\mathbf{v})\rangle(\psi))$ is the translation $\langle \mathcal{T}, g^0(\mathbf{v})\rangle(\varphi)$, we get $\mathfrak{M}, s \vDash \varphi$ iff $\mathcal{T}(\mathfrak{M}), s \vDash \langle \mathcal{T}, g^0(\mathbf{v})\rangle(\varphi)$.

(4). $\varphi = D^n_X x_i$. Then, $\mathfrak{M}, s \vDash \varphi$ iff for all admissible assignment t, $s =_X t$ implies $g^n(s) =_{x_i} g^n(t)$. In FOL terms, this says exactly that $\forall \mathbf{z}\forall \mathbf{z}'(A\mathbf{v} \wedge A\mathbf{v}[\mathbf{z}'/\mathbf{z}] \to g^n_i(\mathbf{v}) = g^n_i(\mathbf{v}[\mathbf{z}'/\mathbf{z}]))$, where $\mathbf{z}'$ is the $\mathbb{V}'$-copy corresponding to $\mathbf{z}$. Again, it is easy to see that $\mathfrak{M}, s \vDash \varphi$ iff $\mathcal{T}(\mathfrak{M}), s \vDash \langle \mathcal{T}, g^0(\mathbf{v})\rangle(\varphi)$. ∎

4.2.4 Changing to a modal semantics

Now we switch to a modal perspective on DFD and its models, which will be our main vehicle in what follows. Our dependence quantifiers D_X are essentially modalities for equivalence relations, and this can be made precise as follows:

Definition 4.6: A *standard relational model* is a tuple $\mathcal{M} = \langle W, g, \sim, V\rangle$[1] s.t.

- W is a non-empty set of possible worlds or states.
- $g : W \to W$ is a total function.
- For each variable $x \in \mathbb{V}$, $\sim_x \subseteq W \times W$ is an equivalence relation. For sets of variables $X \subseteq \mathbb{V}$, we set $\sim_X := \bigcap_{x \in X} \sim_x$.
- Function g preserves $\sim_{\mathbb{V}}$, i.e., $s \sim_{\mathbb{V}} t$ implies $g(s) \sim_{\mathbb{V}} g(t)$.
- V is a valuation map from atoms $P\mathbf{x}$ to $\mathcal{P}(W)$ such that, whenever $s \sim_X t$ and $s \in V(Px_1 \cdots x_n)$ for some $x_1, \cdots, x_n \in X$, then $t \in V(Px_1 \cdots x_n)$.

A pair of a model and world $\langle \mathcal{M}, s\rangle$ (for short, $\mathcal{M}, s$) is called a pointed model.

The set W replaces assignments by abstract states, dropping concrete values assigned to variables. States s, t are called 'X-equivalent' if $s \sim_X t$. For any $X \subseteq \mathbb{V}$, $\sim_X (s) := \{t \in W \mid s \sim_X t\}$, where we write $\sim_x (s)$ for $\sim_{\{x\}} (s)$. Also, for any sets W_1, W_2 of states, we write $W_1 \sim_X W_2$ when $s \sim_X t$ for all $s \in W_1$ and $t \in W_2$. Now we turn to our semantics on standard relational models.

Definition 4.7: Given a pointed model $\langle \mathcal{M}, s\rangle$ and a formula $\varphi \in \mathcal{L}_{\mathsf{D}}$, the following recursion defines when *φ is true in $\mathcal{M}$ at s*, written $\mathcal{M}, s \vDash \varphi$ (where we suppress the clauses that read exactly as in Definition 4.3):

$$\begin{aligned}
\mathcal{M}, s \vDash P\mathbf{x} \quad &\text{iff} \quad s \in V(P\mathbf{x}) \\
\mathcal{M}, s \vDash D^n_X y \quad &\text{iff} \quad \text{for each } t \in W,\ s \sim_X t \text{ implies } g^n(s) \sim_y g^n(t) \\
\mathcal{M}, s \vDash \mathsf{D}_X \varphi \quad &\text{iff} \quad \text{for each } t \in W,\ s \sim_X t \text{ implies } \mathcal{M}, t \vDash \varphi
\end{aligned}$$

1 We use the same notation as for dynamic dependence models $\mathfrak{M}$ to stress the analogy.

Given a model $\mathcal{M}$ and a formula φ, $[\![\varphi]\!]^{\mathcal{M}} := \{t \in W \mid \mathcal{M}, t \vDash \varphi\}$ is the 'truth set' of φ in $\mathcal{M}$, where the index $\mathcal{M}$ is dropped when the model is understood.

4.2.5 Equivalence of the two semantics

We now relate the two semantics by two transformations.

Definition 4.8: For each dynamic dependence model $\mathfrak{M} = \langle O, I, A, g\rangle$, the *induced standard relational model* is $\mathcal{M}^{\mathfrak{M}} = \langle W^{\mathfrak{M}}, g^{\mathfrak{M}}, \sim^{\mathfrak{M}}, V^{\mathfrak{M}}\rangle$ with

- $W^{\mathfrak{M}} := A$
- $g^{\mathfrak{M}} := g$
- For all $s, t \in W^{\mathfrak{M}}$ and $x \in \mathbb{V}$, $s \sim_x^{\mathfrak{M}} t$ iff $s =_x t$
- For each $P\mathbf{x}$, $V^{\mathfrak{M}}(P\mathbf{x}) := \{s \in A \mid \mathfrak{M}, s \vDash P\mathbf{x}\}$.

Here the function $g^{\mathfrak{M}}$ and the valuation function $V^{\mathfrak{M}}$ satisfy the two special conditions imposed in Definition 4.6. In particular, the transition function, being defined on assignments, will give the same values on two $\sim_{\mathbb{V}}$-related assignments since these are identical. Now, a simple induction on formulas φ suffices to show that the FOL semantics agrees with the modal semantics:

Proposition 4.3: For each dynamic dependence model $\mathfrak{M}$ and $\varphi \in \mathcal{L}_{\mathsf{D}}$,

$$\mathfrak{M}, s \vDash \varphi \Leftrightarrow \mathcal{M}^{\mathfrak{M}}, s \vDash \varphi.$$

Here is the, less obvious, transformation in the opposite direction.

Definition 4.9: For each standard relational model $\mathcal{M} = \langle W, g, \sim, V\rangle$ the *induced dynamic dependence model* $\mathfrak{M}^{\mathcal{M}} = \langle O^{\mathcal{M}}, I^{\mathcal{M}}, A^{\mathcal{M}}, g^{\mathcal{M}}\rangle$ has

- $O^{\mathcal{M}} := \{\sim_x (s) \mid s \in W, x \in \mathbb{V}\}$.
- For each n-ary predicate symbol P, $I^{\mathcal{M}}(P) := \{\langle \sim_{x_1} (s), \cdots, \sim_{x_n} (s)\rangle \mid x_{1 \leqslant i \leqslant n} \in \mathbb{V}$ *and* $s \in W$ *with* $s \in V(Px_1 \cdots x_n)\}$.
- $A^{\mathcal{M}} := \{s^{\sim} \mid s \in W\}$ with $s^{\sim}(x) = \sim_x (s)$ for all variables $x \in \mathbb{V}$.
- $g^{\mathcal{M}}(s^{\sim}) = t^{\sim}$ if there are $s', t' \in W$ such that $s \sim_{\mathbb{V}} s'$, $t \sim_{\mathbb{V}} t'$ and $g(s') = t'$.

Here the last clause makes sure that the transition function defined in this way does not depend on the particular representative of the equivalence class.

Again the semantics are in harmony:

Proposition 4.4: For each relational $\mathcal{M} = \langle W, g, \sim, V \rangle$ and $\varphi \in \mathcal{L}_{\mathsf{D}}$, it holds that

$$\mathcal{M}, s \vDash \varphi \Leftrightarrow \mathfrak{M}^{\mathcal{M}}, s^{\sim} \vDash \varphi.$$

Proof We use induction on φ. The cases for atoms and Boolean connectives are routine. The equivalence for atoms $D^n_X y$ holds by the semantics and the following fact, for all $s, t \in W$, $X \subseteq \mathbb{V}$ and $n \in \mathbb{N}$:

$$g^n(s) \sim_X g^n(t) \text{ iff } (g^{\mathcal{M}})^n(s^{\sim}) =_X (g^{\mathcal{M}})^n(t^{\sim})$$

The inductive cases for $\mathsf{D}_X\varphi$ and $\bigcirc\varphi$ are straightforward. ∎

Propositions 4.3 and Proposition 4.4 immediately imply a validity reduction both ways:

Proposition 4.5: The same $\mathcal{L}_{\mathsf{D}}$-formulas are valid on dynamic dependence models and on standard relational models.

Both perspectives on the logic DFD are interesting, but we will mainly work with the modal view, which allows us to use notions such as generated submodels, bisimulations, and *p*-morphisms, and techniques such as unraveling (see, e.g., Blackburn et al., 2001).

4.3 Axiomatizing DFD and consideration on its complexity

The first-order translation of Section 4.2.3 shows that the validities of DFD are effectively axiomatizable, since they are for FOL, van Benthem (1984). Now we present a concrete complete Hilbert-style proof calculus. After that, we show some basic result on the decidability of a significant fragment of DFD.

4.3.1 The proof system **DFD**

The proof system **DFD** for dynamic dependence logic is presented in Table 4.1. The notions of syntactical derivation and provability are defined as usual.

Here, axioms ○-Distribution and D-Distribution are standard for normal modalities, the Functionality axiom ensures that dynamic transitions between states are a function. The dependence quantifiers are **S5**-modalities. Note also how Dyn-Trans combines dependence formulas $D^n_X Y$ with the temporal operator ○ to express dependencies over time. Determinism says that fixing the current assignment fixes the values of each variable at

Table 4.1 The proof system **DFD**

I	**Axioms and Rules of Classical Propositional Logic**
II	**Axioms and Rules for $\bigcirc$**
$\bigcirc$-Distribution	$\bigcirc(\varphi \to \psi) \to (\bigcirc\varphi \to \bigcirc\psi)$
Functionality	$\bigcirc\neg\varphi \leftrightarrow \neg \bigcirc \varphi$
$\bigcirc$-Necessitation	From φ, infer $\bigcirc\varphi$
III	**Axioms and Rules for D**
D-Distribution	$\mathsf{D}_X(\varphi \to \psi) \to (\mathsf{D}_X\varphi \to \mathsf{D}_X\psi)$
D-Introduction$_1$	$Px_1 \cdots x_n \to \mathsf{D}_{\{x_1,\cdots,x_n\}} Px_1 \cdots x_n$
D-Introduction$_2$	$D^n_X y \to \mathsf{D}_X D^n_X y$
D-T	$\mathsf{D}_X\varphi \to \varphi$
D-4	$\mathsf{D}_X\varphi \to \mathsf{D}_X\mathsf{D}_X\varphi$
D-5	$\neg\mathsf{D}_X\varphi \to \mathsf{D}_X\neg\mathsf{D}_X\varphi$
D-Necessitation	From φ, infer $\mathsf{D}_X\varphi$
IV	**Axioms for $D^n_X y$**
Dep-Ref	$D_X x$ for all $x \in X$
Dyn-Trans	$D^n_X Y \wedge \bigcirc^n D^m_Y Z \to D^{m+n}_X Z$
Determinism	$D^1_{\mathbb{V}} x$ for all $x \in \mathbb{V}$
V	**Interaction Axioms**
Transfer	$D^n_X Y \wedge \bigcirc^n \mathsf{D}_Y\varphi \to \mathsf{D}_X \bigcirc^n \varphi$
D-$\bigcirc$	$\mathsf{D}_\varnothing\varphi \to \bigcirc\varphi$

the next stage: i.e., transitions only depend on global system states, not on when these states occur. Finally, D-$\bigcirc$ says that universal truth implies what is true in the future.

The system **DFD** can prove some interesting theorems. For instance,

- $D^n_X Y \wedge D^n_Z U \to D^n_{X\cup Z}(Y \cup U)$ (Additivity of Dynamic Dependence)
- $D^n_X y \to D^n_Z y$, for $X \subseteq Z$ (Monotonicity of Dynamic Dependence)
- $\mathsf{D}_X \varphi \to \mathsf{D}_Y \varphi$, for $X \subseteq Y$ (Monotonicity of Dependence Quantifiers)
- $\bigcirc \mathsf{D}_{\mathbb{V}} \varphi \to \mathsf{D}_{\mathbb{V}} \bigcirc \varphi$ ($\bigcirc$-$\mathsf{D}_{\mathbb{V}}$-Commutation)

All these derivable principles are valid.

Proposition 4.6: The calculus **DFD** is sound w.r.t. standard relational models.

Proof We only consider two not entirely routine cases. Let $\mathcal{M} = \langle W, g, \sim, V\rangle$ be a standard relational model and $s \in W$. We show $D^n_X Y \wedge \bigcirc^n D^m_Y Z \to D^{m+n}_X Z$ and $D^n_X Y \wedge \bigcirc^n \mathsf{D}_Y \varphi \to \mathsf{D}_X \bigcirc^n \varphi$ are true.

Let $\mathcal{M}, s \vDash D^n_X Y \wedge \bigcirc^n D^m_Y Z$, and consider any $t \in W$ with $s \sim_X t$. From $D^n_X Y$ at s, it follows that $g^n(s) \sim_Y g^n(t)$ (a), and since $\bigcirc^n D^m_Y Z$ at s, $D^m_Y Z$ holds at $g^n(s)$ (b). Combining (a) and (b), we get $g^{m+n}(s) \sim_Z g^{m+n}(t)$.

Next, let $\mathcal{M}, s \vDash D^n_X Y \wedge \bigcirc^n \mathsf{D}_Y \varphi$, and consider t with $s \sim_X t$. Again by $D^n_X Y$ at s, we have $g^n(s) \sim_Y g^n(t)$ (a). Moreover, by $\bigcirc^n \mathsf{D}_Y \varphi$ at s, $\mathsf{D}_Y \varphi$ is true at $g^n(s)$ (b). Combining (a) and (b), φ is true at $g^n(t)$, i.e., $\bigcirc^n \varphi$ is true at t. ∎

By Proposition 4.5 then, **DFD** is sound for dynamic dependence models too.

4.3.2 Completeness of **DFD**: introducing general relational models

Our eventual aim is to show that the system **DFD** is complete w.r.t. standard relational models. To achieve this, we take three steps of separate interest:

Step 1. We introduce a new notion of 'general relational models' and interpret $\mathcal{L}_{\mathsf{D}}$-formulas in this broader setting.

Step 2. We prove completeness of the system **DFD** w.r.t. the new models.

Step 3. We prove a representation result for general relational models as p-morphic images of standard relational models, which implies that **DFD** is also complete w.r.t. standard relational models.

This subsection is concerned with Step 1.

Definition 4.10: A *general relational model* is a tuple $\mathcal{M}_g = \langle A, g, =_X, D^n_X y, P\mathbf{x}\rangle$ with the following components:

- A is a non-empty set of possible worlds or states.
- $g : A \to A$ is a total function.
- For each set $X \subseteq \mathbb{V}$ of variables, $=_X \subseteq A \times A$ is a binary relation.
- For each n-ary predicate symbol P and each world $s \in A$, P^s is an n-ary relation on variables.
- For each $s \in A$, $(D^n_X y)^s \subseteq \mathcal{P}(\mathbb{V}) \times \mathbb{V}$ is a relation between finite sets of variables X and variables y.

These ingredients are required to satisfy the following conditions:

C1 All $=_X$ are equivalence relations on A, and $=_\emptyset$ is the universal relation

C2 All $(D^n_X y)^s$ satisfy the following three properties:

- Dep-Reflexivity: For all $x \in X \subseteq \mathbb{V}$, it holds that $(D_X x)^s$.
- Dyn-Transitivity: If $(D^n_X Y)^s$ and $(D^m_Y Z)^{g^n(s)}$, then $(D^{n+m}_X Z)^s$.
- Determinism: For all $x \in \mathbb{V}$, $(D^1_{\mathbb{V}} x)^s$.

C3 If $s =_X t$ and $(D^n_X Y)^s$, then $(D^n_X Y)^t$ and $g^n(s) =_Y g^n(t)$

C4 If $s =_X t$, $(P\mathbf{y})^{g^n(s)}$ and $(D^n_X Y)^s$ (where Y is the set of variables occurring in the sequence $\mathbf{y}$), then $(P\mathbf{y})^{g^n(t)}$.

It is useful to compare this new notion with the standard relational models of Definition 4.6. Each $=_X$ for $X \subseteq \mathbb{V}$ is now a primitive relation, not necessarily the intersection of the individual $=_{x \in X}$. Also, for each $n \in \mathbb{N}$, formulas $D^n_X y$ are treated as atoms now, with truth values given directly by valuation functions.

The resulting truth definition for $\mathcal{L}_{\mathsf{D}}$-formulas in general relational models reads exactly as that in Definition 4.7 for standard relational models, though with the new understanding of relations and atoms as just explained.

Proposition 4.7: The proof system **DFD** is sound for general relational models.

Proof Condition C1 guarantees the validity of the standard modal principles of the calculus. Condition C2 gives us Dep-Ref, Dyn-Trans and Determinism. Finally, condition C3 gives both D-Introduction$_2$ and Transfer. ∎

General relational models are not just a stepping stone toward the completeness theorem for DFD. They also have an independent interest as modal semantic structures. To illustrate this, we show how the standard notion of bisimulation applies, of which we will use the special case of p-morphisms later on.

Definition 4.11: Consider two general relational models $\mathcal{M}_g = \langle A, g, =_X, D^n_X y, P\mathbf{x}\rangle$, $\mathcal{M}'_g = \langle A', g', ='_X, D^n_X y, P\mathbf{x}\rangle$ with $s \in A$, $s' \in A'$ and $X \cup \{y\} \subseteq \mathbb{V}$. A DFD-*bisimulation* is a relation Z_{D} between pointed models s.t., if $\langle \mathcal{M}, s\rangle Z_{\mathsf{D}} \langle \mathcal{M}', s'\rangle$, then

Atom For both types of atoms α (factual $P\mathbf{x}$ and dependence-style $D^n_X y$), $\mathcal{M}_g, s \vDash \alpha$ iff $\mathcal{M}'_g, s' \vDash \alpha$

Func $\langle \mathcal{M}_g, g(s)\rangle \; Z_{\mathsf{D}} \; \langle \mathcal{M}'_g, g'(s')\rangle$

Forth$_{\mathsf{D}}$ If $s =_X t$, then there is a $t' \in W'$ s.t. $s' ='_X t'$ and $\langle \mathcal{M}_g, t\rangle \; Z_{\mathsf{D}} \; \langle \mathcal{M}'_g, t'\rangle$

Back$_{\mathsf{D}}$ If $s' ='_X t'$, then there is a $t \in W$ s.t. $s =_X t$ and $\langle \mathcal{M}_g, t\rangle \; Z_{\mathsf{D}} \; \langle \mathcal{M}'_g, t'\rangle$.

For a set $B \subseteq A$ of states, $g^n(B) := \{g^n(s) \mid s \in B\}$. Finally, we also write $\langle \mathcal{M}, s\rangle \underline{\leftrightarrow}_{\mathsf{D}} \langle \mathcal{M}', s'\rangle$ if $\langle \mathcal{M}, s\rangle \; Z_{\mathsf{D}} \; \langle \mathcal{M}', s'\rangle$ for some DFD-bisimulation Z_{D}.

Here, Clauses **Atom**, **Forth$_{\mathsf{D}}$** and **Back$_{\mathsf{D}}$** are standard for bisimulation (Blackburn et al., 2001). **Func** handles state transitions. Also, while we imposed just equivalence of truth values for dynamic dependence atoms $D^n_X y$, this requirement unpacks into a more complex constraint on accessibility relations in standard relational models.

We say that two pointed models $\langle \mathcal{M}_g, s\rangle$ and $\langle \mathcal{M}'_g, s'\rangle$ are DFD-*equivalent* (written as $\langle \mathcal{M}_g, s\rangle \leftrightsquigarrow_{\mathsf{D}} \langle \mathcal{M}'_g, s'\rangle$) if, for all formulas $\varphi \in \mathcal{L}_{\mathsf{D}}$, $\mathcal{M}_g, s \vDash \varphi$ iff $\mathcal{M}'_g, s' \vDash \varphi$.

Proposition 4.8: $\langle \mathcal{M}_g, s\rangle \underline{\leftrightarrow}_{\mathsf{D}} \langle \mathcal{M}'_g, s'\rangle$ implies $\langle \mathcal{M}_g, s\rangle \leftrightsquigarrow_{\mathsf{D}} \langle \mathcal{M}_g, s'\rangle$.

The proof is a straightforward induction on formulas. Also easy to prove is a partial converse, viz. a Hennessy-Milner Theorem for *image-finite* relational model $\mathcal{M}_g = \langle A, g, =_X, D^n_X y, P\mathbf{x}\rangle$ where, for all $s \in A$ and $X \subseteq \mathbb{V}$, $\{t \in A \mid s =_X t\}$ is finite.

Proposition 4.9: For any two image-finite pointed general relational models, it holds that $\langle \mathcal{M}_g, s\rangle \leftrightsquigarrow_{\mathsf{D}} \langle \mathcal{M}'_g, s'\rangle$ implies that $\langle \mathcal{M}_g, s\rangle \underline{\leftrightarrow}_{\mathsf{D}} \langle \mathcal{M}'_g, s'\rangle$.

These are just a few illustrations. A wide range of modal techniques (cf. Blackburn et al., 2001) for the basic theory, applies to general relational models for DFD.

4.3.3 Canonical models for DFD

We now come to Step 2. Showing that the system **DFD** is complete w.r.t. general relational models appeals to a standard construction in modal logic.

Definition 4.12: The *canonical model for* **DFD** is the structure $\mathcal{M}^c = \langle W^c, g^c, =^c_X, V^c\rangle$, where

- W^c is the class of all maximal **DFD**-consistent sets

- For all $s \in W^c$, $g^c(s) = \{\varphi \in \mathcal{L}_{\mathsf{D}} \mid \bigcirc\varphi \in s\}$
- For all $s, t \in W^c$ and $X \subseteq \mathbb{V}$, $s =^c_X t$ iff $\mathsf{D}_X s \subseteq t$
- $s \in V^c(D^n_X y)$ iff $D^n_X y \in s$, and $s \in V^c(P\mathbf{x})$ iff $P\mathbf{x} \in s$.

For all states $s \in W^c$, $\mathsf{D}_X s$ denotes the set of formulas $\{\varphi \mid \mathsf{D}_X\varphi \in s\}$.

That g^c defines a function on W^c follows from this observation:

Proposition 4.10: In the canonical model $\mathcal{M}^c = \langle W^c, g^c, =^c_X, V^c\rangle$, $g^c(s) \in W^c$.

Proof Since the Functionality axiom of **DFD** has syntactic Sahlqvist form, the canonical model will satisfy its corresponding semantic frame condition of functionality by a standard argument about maximally consistent sets (Blackburn et al., 2001). ∎

So, the similarity type of the model fits. It remains to check the conditions on general relational models listed in Definition 4.10.

Remark 4.1: Before proceeding, we must address a small problem first, viz. the fact that the relation $=^c_\emptyset$ as defined earlier in the canonical model for **DFD** need not be the real *universal relation* in that model.

To get around this, we use a standard technique from completeness proofs for modal logics containing a global universal modality (Goranko and Passy, 1992). Instead of taking the whole canonical model introduced above, we start from any world $u \in W^c$ and restrict the states to those in the *generated submodel* in the relation $=^c_\emptyset$. Then, the proof principles of the calculus **DFD** guarantee that the accessibility relations for the other dependence and temporal modalities are contained in $=^c_\emptyset$, and thus, we have all essential structure available within the generated submodel.

With this understanding, when we talk about the canonical model in what follows, we really mean any generated submodel of the sort described.

Proposition 4.11: The canonical model is a general relational model.

Proof (1). Given the **S5**-axioms for dependence quantifiers, all $=^c_X$ are equivalence relations by a standard modal argument. Also, since we now talk about generated canonical models, the relation $=^c_\emptyset$ is the universal relation.

(2). The **DFD** axioms for $D^n_X y$ were precisely designed to ensure the truth of the conditions of 'Dep-Reflexivity', 'Dyn-Transitivity' and 'Determinism'.

(3). Condition C3. Let $s =^c_X t$ and $D^n_X Y \in s$. Using the axiom $D^n_X Y \to \mathsf{D}_X D^n_X Y$, we get $\mathsf{D}_X D^n_X Y \in s$. Since $s =^c_X t$, we have $D^n_X Y \in t$. Next, let $\mathsf{D}_y \varphi \in (g^c)^n(s)$: we show that $\varphi \in (g^c)^n(t)$. By the definition of g^c, $\mathsf{D}_Y \varphi \in (g^c)^n(s)$ implies $\bigcirc^n \mathsf{D}_Y \varphi \in s$. Then, using the axiom $D^n_X Y \wedge \bigcirc^n \mathsf{D}_Y \varphi \to \mathsf{D}_X \bigcirc^n \varphi$, we get $\mathsf{D}_X \bigcirc^n \varphi \in s$, and hence $\bigcirc^n \varphi \in t$. Therefore, $\varphi \in (g^c)^n(t)$.

(4). Condition C4. Let $s =^c_X t$, $P\mathbf{y} \in g^{cn}(s)$ and $D^n_X Y \in s$ (where Y is the set of all variables occurring in the tuple $\mathbf{y}$). We show that $P\mathbf{y} \in g^{cn}(t)$. Using the axiom $Px_1 \cdots x_n \to \mathsf{D}_{\{x_1,\cdots,x_n\}} Px_1 \cdots x_n$, we have $\mathsf{D}_Y P\mathbf{y} \in g^{cn}(s)$, and hence $\bigcirc^n \mathsf{D}_Y P\mathbf{y} \in s$. Now, from $D^n_X Y \wedge \bigcirc^n \mathsf{D}_Y \varphi \to \mathsf{D}_X \bigcirc^n \varphi$, we have $\mathsf{D}_X \bigcirc^n P\mathbf{y} \in s$. Also, as $s =^c_X t$, $\bigcirc^n P\mathbf{y} \in t$. Therefore, $P\mathbf{y} \in g^{cn}(t)$. ∎

Next, a standard argument proves the following Existence Lemma:

Lemma 4.1: Let $\mathcal{M}^c$ be the canonical model and $s \in W^c$. Then we have:

If $\widehat{\mathsf{D}}_X \varphi \in s$, then there exists $t \in W^c$ such that $s =^c_X t$ and $\varphi \in t$.

Now we are able to prove the following key *Truth Lemma*:

Lemma 4.2: Let $\mathcal{M}^c$ be the canonical model, $s \in W^c$ and $\varphi \in \mathcal{L}_{\mathsf{D}}$. Then

$$\mathcal{M}^c, s \vDash \varphi \Leftrightarrow \varphi \in s.$$

Proof The proof is by induction on φ. We only show two cases.

(1). Formula φ is $\bigcirc\psi$. From the semantics, $\mathcal{M}^c, s \vDash \varphi$ iff $\mathcal{M}^c, g^c(s) \vDash \psi$. Then, by the inductive hypothesis, $\mathcal{M}^c, g^c(s) \vDash \psi$ iff $\psi \in g^c(s)$. From the definition of g^c, we know that $\psi \in g^c(s)$ iff $\varphi \in s$.

(2). Formula φ is $\widehat{\mathsf{D}}_X \psi$. From left to right, assume that $\mathcal{M}^c, s \vDash \widehat{\mathsf{D}}_X \psi$. Then, there exists $t \in W^c$ such that $s =^c_X t$ and $\mathcal{M}^c, t \vDash \psi$. By the inductive hypothesis, $\psi \in t$. From the definition of $=^c_X$, we have $\widehat{\mathsf{D}}_X \psi \in s$. Conversely, suppose that $\widehat{\mathsf{D}}_X \psi \in s$. Then, by Lemma 4.1, there is a $t \in W^c$ such that $s =^c_X t$ and $\psi \in t$. By the inductive hypothesis, $\mathcal{M}^c, t \vDash \psi$. Therefore, $\mathcal{M}^c, s \vDash \widehat{\mathsf{D}}_X \psi$. ∎

Steps 1 and 2 are now completed, and together yield the following result.

Theorem 4.1: **DFD** is sound and complete for general relational models.

4.3.4 Representation and completeness for standard models

Now we take the final Step 3 in our proof plan, and represent general relational models as standard relational models. We elaborate the tree-style representation sketched in the Appendix to Baltag and van Benthem (2021b), adapted to our richer temporal setting.

Theorem 4.2: Each general relational model is a p-morphic image of some standard relational model.

Proof Let $\mathcal{M}_g = \langle A, g, =_X, D^n_X y, P\mathbf{x}\rangle$ be a general relational model. To construct a standard relational model $\mathcal{M}^{st}$ linked to $\mathcal{M}_g$, we need worlds W^{st}, a transition function g^{st}, accessibility relations $\sim^{st}_X$, and a suitable valuation V^{st}.

For the *worlds* W^{st}, we take the set of all *histories*, i.e., all finite sequences $h = \langle s_0, \bigtriangleup_1, s_1, \cdots, \bigtriangleup_n, s_n\rangle$ with states from the model $\mathcal{M}_g$ such that:

- For each $i \leqslant n$, $s_i \in A$, and $\bigtriangleup_i$ equals $X_i \subseteq \mathbb{V}$ or g.
- For each k s.t. $1 \leqslant k \leqslant n$, if $\bigtriangleup_k$ is X_k, then $s_{k-1} =_{X_k} s_k$; and if $\bigtriangleup_k$ is g, then $g(s_{k-1}) = s_k$.

These finite histories from a natural tree- or forest-like structure. In particular, any two histories h, h' that share the same initial state are connected by a unique path that can be pictured in terms of first 'going down' from h to the largest shared sub-history, and then 'going up' again to the end of h'. This visual picture may help in understanding the arguments to follow.

Let $last(h)$ be the last state in history h. For the *transition function*, we set:

- $g^{st}(h) = \langle h, g, g(last(h))\rangle$

Next, we define the *valuation* V^{st} for atoms $P\mathbf{x}$ simply in terms of truth at the last world in the history:

- $h \in V^{st}(P\mathbf{x})$ iff $\mathcal{M}_g, last(h) \vDash P\mathbf{x}$.

The final, and most delicate task is to define the *accessibility relations* in the model $\mathcal{M}^{st}$. This has to be done in such a way that we 'improve' the given general relational model in two respects: (i) relations $\sim_X$ become intersections of the $\sim_x$ for $x \in X$, (ii) atoms $D_X y$ get their standard semantic interpretation at histories h in a way that matches with their truth in $\mathcal{M}_g$ at $last(h)$.

First, we introduce relations $h \rightsquigarrow_X h'$ between histories as follows:

- h is of the form $\langle h_0, g, s_1, g, s_2, \cdots, g, s_n\rangle$ and for some $Y \subseteq \mathbb{V}$ with $\mathcal{M}_g, last(h_0) \vDash D^n_Y X$, $h' = \langle h_0, Y, last(h'_0), g, t_1, g, t_2, \cdots, g, t_n\rangle$

Here, we use the notation $last(h_0')$ to highlight that it is the last state in the initial segment of h' before the final n-step action of the transition function.

In the extreme case $n = 0$, there are no transition steps, and $h \rightsquigarrow_X h'$ just says that history h' is of the form $\langle h, Y, last(h')\rangle$ with $\mathcal{M}_g, last(h) \vDash D_Y X$. However, one should note, for later reference, that the dependence atom $D_Y^n X$ present at the start 'manifests' itself as an X-dependence between longer histories that have a further tail of n consecutive transition steps.

The one-step relation $\rightsquigarrow_X$ need not be an equivalence relation, as is needed, and therefore, the actual *accessibility relation of* $\mathcal{M}^{st}$ is defined as follows:

- $\leftrightsquigarrow_X$ is the *reflexive-transitive-symmetric closure* of $\rightsquigarrow_X$.

This completes our definition of the model $\mathcal{M}^{st}$ that we are going to use.

Now we must prove two basic facts that, together, establish our theorem.

Lemma 4.3: $\mathcal{M}^{st} = \langle W^{st}, g^{st}, \leftrightsquigarrow_X, V^{st}\rangle$ is a standard relational model.

The proofs to follow should be understood as concentrating on the essentials. The accessibility relation in $\mathcal{M}^{st}$ is defined as being connected by some finite sequence of the above basic steps (in either order). Proving facts about this notion can be done by natural induction, but instead of going through this routine, wherever possible, we will explain the case of the single steps, since usually, this behavior lifts automatically to the whole sequence.

Proof The first condition in the surplus of standard relational models over general relational models is that the relation $\leftrightsquigarrow_X$ equals the intersection $\bigcap_{x\in X} \leftrightsquigarrow_x$. This is a crucial feature of the above tree construction, which cannot be enforced routinely by means of the standard accessibility relations in canonical models (Gargov and Passy, 1990). We therefore state it as a separate fact.

Claim 1. For any two histories h and h', $h \leftrightsquigarrow_X h'$ iff $h \leftrightsquigarrow_x h'$ for all $x \in X$.

Proof It suffices to show this property for single steps in the relation $h \leftrightsquigarrow_X h'$ since it will transfer automatically to longer sequences. The essential observations are these. Suppose we have a transition with an initial Y-step in the history h' and $D_Y^n X$ true at $last(h_0)$ in $\mathcal{M}_g$. By the Dep-Reflexivity assumption on general relational models, for any $x \in X$, $\bigcirc^n D_X x$ is true at $last(h_0)$. Then, by the Dyn-Transitivity, $D_Y^n x$ will be true at $last(h_0)$, and so $h \leftrightsquigarrow_x h'$. Conversely, if we have a single initial Y-step that qualifies for all transitions for the variables $x \in X$, then we have $D_Y^n x$ true at $last(h_0)$ in $\mathcal{M}_g$, and by

the convention in Definition 4.1, this means that $D^n_Y X$ is true there: so we also have an $\leftrightsquigarrow_X$ step. ∎

Next, we show that the transition function g^{st} respects $\leftrightsquigarrow_{\mathbb{V}}$.

Claim 2. If $h \leftrightsquigarrow_{\mathbb{V}} h'$, then $g^{st}(h) \leftrightsquigarrow_{\mathbb{V}} g^{st}(h')$.

Proof Again it suffices to analyze single steps for this type of assertion. We only consider the case with $h \rightsquigarrow_{\mathbb{V}} h'$, that of $h' \rightsquigarrow_{\mathbb{V}} h$ is similar. We have $h'_0 = \langle h_0, Z, u\rangle$ with $last(h_0) \in V(D^n_Z \mathbb{V})$. Also, by the Determinism property in Definition 4.10, $last(h) \in V(D^1_{\mathbb{V}} \mathbb{V})$. Therefore, by our definition of histories and our semantics on general relational models, $\mathcal{M}_g, last(h_0) \vDash \bigcirc^n D^1_{\mathbb{V}} \mathbb{V}$. Then, by the Dyn-Transitivity of Definition 4.10 in $\mathcal{M}_g$, $last(h_0) \in V(D^{n+1}_Z \mathbb{V})$. It follows that $g^{st}(h) \rightsquigarrow_{\mathbb{V}} g^{st}(h')$, and therefore $g^{st}(h) \leftrightsquigarrow_{\mathbb{V}} g^{st}(h')$. ∎

The third condition to be shown is that the truth values of non-dependence atoms are invariant in the way required by Definition 4.6.

Claim 3. If $h \leftrightsquigarrow_X h'$ and $h \in V^{st}(Px_1 \cdots x_k)$ for some $x_1, \cdots, x_k \in X$, then also $h' \in V^{st}(Px_1 \cdots x_k)$.

Proof Once more, it suffices to prove the claim by analyzing a single step transition, and this time, we illustrate both cases $h \rightsquigarrow_X h'$ and $h' \rightsquigarrow_X h$.

Case 1: $h \rightsquigarrow_X h'$, i.e., $h'_0 = \langle h_0, Z, u\rangle$ for some Z and u, and $last(h_0) \in V(D^k_Z X)$. From the assumption that $h \in V^{st}(Px_1 \cdots x_k)$, reasoning as above for Claim 2, we have that $\mathcal{M}_g, last(h_0) \vDash \bigcirc^n Px_1 \cdots x_k$. Combining this with $last(h_0) \in V(D^k_Z X)$, and using closure property C4 of general relational models, we conclude that $g^n(u) \vDash Px_1 \cdots x_k$, i.e., $\mathcal{M}_g, last(h) \vDash Px_1 \cdots x_k$.

Case 2: $h' \rightsquigarrow_X h$. This time, $h_0 = \langle h'_0, Z, u\rangle$ for some Z and u s.t. $last(h'_0) \in V^{st}(D^k_Z X)$ and $g^k(u) \in V^{st}(Px_1 \cdots x_k)$. Now, from $last(h'_0) =_Z u$ and $last(h'_0) \in V^{st}(D^k_Z X)$ we get $u \in V^{st}(D^k_Z X)$, by C3 in Definition 4.10. Finally, by clause C4 in Definition 4.10, $last(h'_0) =_Z u$, $u \in V^{st}(D^k_Z X)$ and $g^k(u) \in V^{st}(Px_1 \cdots x_k)$ imply $g^k(last(h'_0)) \in V^{st}(Px_1 \cdots x_k)$, i.e., $h' \in V^{st}(Px_1 \cdots x_k)$. ∎

This completes the proof of Lemma 4.3. ∎

Finally, we define a map F from our model $\mathcal{M}^{st}$ to the original general relational model $\mathcal{M}_g$. We simply put, for all histories $h \in W^{st}$:

- $F(h) = last(h)$

What remains is to check that this map is a functional version of the bisimulations introduced in Definition 4.11 (in modal terminology, it is a 'p-morphism'), which preserve the truth values of formulas in the language of DFD.

Lemma 4.4: The function F is a modal p-morphism from $\mathcal{M}^{st}$ onto $\mathcal{M}_g$.

Proof First, surjectivity is obvious, since each $s \in A$ equals the function value $F(\langle s \rangle)$. Next, and much less straightforwardly, we must check that the map F satisfies the back-and-forth clauses of modal p-morphisms for the dependence relations and for the transition function, as well as the 'harmony' clause for the two kinds of atoms. We state these with their reasons.

- If $h \leftrightsquigarrow_X h'$, then $last(h) =_X last(h')$.
  It suffices to look at single steps, and one case will show why the assertion is true. Suppose that $h' \rightsquigarrow_X h$. Then $h_0 = \langle h'_0, Z, u \rangle$ and $last(h'_0) \in V(D^n_Z X)$. As $last(h'_0) =_Z u$, the condition C3 in Definition 4.10 for general models gives us that $g^n(last(h'_0)) =_X g^n(u)$, i.e., $last(h') =_X last(h)$.
- If $last(h) =_X s$, then there is a history h' with $h \leftrightsquigarrow_X h'$ and $last(h') = s$.
  For h', we can just take the history $\langle h, X, s \rangle$.
- If $g^{st}(h) = h'$, then $g(last(h)) = last(h')$.
  Since $h' = \langle h, g, g(last(h)) \rangle$, this is true by the definition of g^{st}.
- If $g(last(h)) = s$, then there is a history h' with $g^{st}(h) = h'$ and $last(h') = s$.
  Here it suffices to let h' be the history $\langle h, g, s \rangle$.

Next, we consider the valuation on atoms. For standard atoms $P\mathbf{x}$, histories h in $\mathcal{M}^{st}$ agree with their F-values $last(h)$ in $\mathcal{M}_g$ by the definition of V^{st}.

The more challenging case is that of dependence atoms, since these get their meaning through the semantics in the standard relational model $\mathcal{M}^{st}$ rather then being imposed by the valuation. Thus, we need to show the following:

- $\mathcal{M}^{st}, h \vDash D^n_X y$ iff $\mathcal{M}_g, last(h) \vDash D^n_X y$

Proof We first make the auxiliary observation that local dependence statements are preserved along sequences of single steps for the relation $\leftrightsquigarrow_X$.

Claim 4. If $h \leftrightsquigarrow_X h'$ and $\mathcal{M}_g, last(h) \vDash D^n_X y$, then $\mathcal{M}_g, last(h') \vDash D^n_X y$.

The proof is a simple application of Property C3 of general relational models, used in the same way as in the proofs of Claims 2 and 3 above.

Next we spell out the fact about dependence atoms that was needed above.

Claim 5. The following two assertions are equivalent for histories h:

(a) $h \leftrightsquigarrow_X h'$ implies $(g^{st})^n(h) \leftrightsquigarrow_y (g^{st})^n(h')$ for all histories h'

(b) $\mathcal{M}_g, last(h) \vDash D^n_X y$.

Proof *From (a) to (b).* Let $last(h) := s$ and define h' to be the history $\langle h, X, s\rangle$ which is one step longer than h. By our definitions, $h \leftrightsquigarrow_X h'$, and therefore, by our assumption, $(g^{st})^n(h) \leftrightsquigarrow_y (g^{st})^n(h')$. Clearly, the length difference between h and h' just persists after the added n transition steps toward their g^n-values. Given this, the $\leftrightsquigarrow_y$-connection between $(g^{st})^n(h)$ and $(g^{st})^n(h')$ can only have come about by one single $\rightsquigarrow_y$-step, going from the former to the latter history. Moreover, given the definition of these steps, it cannot have occurred in the final parts of these histories, since these have transitions marked by g. The only possibility that remains is that the transition from h to h' was itself a $\rightsquigarrow_y$-step, and by definition, this can only have been because $\mathcal{M}_g, last(h) \vDash D^n_X y$.

From (b) to (a). Let $h \leftrightsquigarrow_X h'$. Again we analyze a single transition step. Let $\mathcal{M}_g, last(h_0) \vDash D^k_Z X$ where k is the number of final g-steps in h, while h' starts with $\langle h_0, Z, u\rangle$. Since $\mathcal{M}_g, last(h) \vDash D^n_X y$, we also have $\mathcal{M}_g, last(h_0) \vDash \bigcirc^k D^n_X y$. But then, using the Dyn-Transitivity of general relational models, we have $\mathcal{M}_g, last(h_0) \vDash D^{n+k}_Z y$, and this implies by definition that $(g^{st})^n(h) \rightsquigarrow_y (g^{st})^n(h')$ and hence also that $(g^{st})^n(h) \leftrightsquigarrow_y (g^{st})^n(h')$. ∎

Taking all this together, F is a surjective p-morphism from $\mathcal{M}^{st}$ to $\mathcal{M}_g$. ∎

This completes the proof of Theorem 4.2. ∎

As truth of modal formulas is preserved under surjective p-morphisms, (Blackburn et al., 2001), and standard relational models are general relational models, it follows that the same $\mathcal{L}_{\mathsf{D}}$-formulas are valid on general relational models and on standard relational models. Combining this with the earlier representation results of Section 4.2.5, we have shown completeness of our proof system in the following sense.

Theorem 4.3: The proof system **DFD** is sound and complete w.r.t. both standard relational models and dynamic dependence models.

4.3.5 Considerations on the decidability of DFD

The next obvious concern would be the complexity of the logic DFD. So far, we have not been able to determine whether the logic DFD is decidable, although we strongly

suspect that it is. However, some partial results exist. In the Appendix to this chapter, we have included a proof that the more 'local' fragment of DFD without operators $\mathsf{D}_{\varnothing}\varphi$ and $D^n_{\varnothing}y$ has the effective finite model property with regard to general relational models[1], and therefore, it is decidable.

4.4 Continuous dependence: the topological logic DCD

Dynamical systems usually come with a topology on their state space, and a continuous transition function between states. We will now extend our basic logic DFD to a new logic DCD that can deal with this richer setting.

4.4.1 Varieties of topological dependence

While the relevant topology on a set of system states can arise in many ways, we will take a particular approach in what follows, and generate them from topologies on the sets of values that the variables of the system can take. This is the proper setting for notions of *approximation* of values, with open sets viewed as possible outcomes of *measurements*. The topology on states can be derived from these value topologies. This allows for more realistic epistemic scenarios where we do not know the function values analyzed in our first dependence logic DFD, but can approximate them to any required degree of precision.

This topologization lays a bridge between dynamic dependence and existing *dynamic topological logics* which encode reasoning about the stepwise action of continuous functions on topological spaces, and in some richer versions, even asymptotic behavior of the dynamical system. Dynamic temporal logics have a temporal next-state modality $\bigcirc$ interpreted as we did in earlier sections, but also a standard topological *interior modality* $\Box\varphi$ true at the interior of the truth set of φ (van Benthem and Bezhanishvili, 2007). This language can express that the transition function is continuous by means of the axiom $\bigcirc\Box\varphi \rightarrow \Box\bigcirc\varphi$ cf. (Kremer and Mints, 2007) which has many more details on guiding ideas and results in dynamic topological logic.

All these ideas return in our logic DCD of *dynamic continuous dependence*, which can be seen as a combination of DTL with our dependence logic DFD.

What needs to happen for this to work is fixing a suitable notion of 'dynamic continuous dependence'. As it happens, the rich topological setting offers different options

1 As the fragment does not contain formulas $\mathsf{D}_{\varnothing}\varphi$ or $D^n_{\varnothing}y$, we need not consider the clause on the universal relation $=_{\varnothing}$ in Definition 4.10.

for this. The expression '*the n-th step value of a variable y depends continuously on the values of X*' may mean at least the following things:

- *Local Version*. For the current state s, $g^n(s)(y)$: the n-th step value of variable y at s is determined to any desired degree of accuracy by some degree of accuracy of the values of the variables X at s.

  However, this notion is 'sensitive' in that it may fail at just slightly different states s of a dynamical system. Here is a more stable version:

- *Global Version*. For all states s, $g^n(s)(y)$ is determined to any desired degree of accuracy by some degree of approximation of the values $s(X)$.

  The stability obtained in this way seems too strong. In a setting of approximation, it will often suffice to have dependence in a suitable zone around the current state. This brings us to our preferred notion for what follows:

- *Neighborhood Version.* There is some desired degree of accuracy U for the actual values $s(X)$ such that for all states t with $t(X) \in U$ (i.e., the values $t(X)$ are close enough to the actual ones at s), $g^n(t)(y)$ is determined to any desired degree of approximation by some degree of approximation of $t(X)$.

In epistemic terms, the local version is a dependency that might be unknowable to the agent in an empirical setting with only approximate measurements. The global version expresses a dependency that is actually known, whereas the neighborhood version describes a dependency that is *knowable*: it might become known by learning enough about the current state of the system.[1]

We will work with the knowability version of topological dependence in what follows, but there are other natural options, too: see Baltag and van Benthem (2021a) for a broader landscape.

4.5 The logic DCD: language and semantics

We now start defining our logic DCD. Its *language* $\mathcal{L}_D$ is the same as the earlier $\mathcal{L}_{\mathsf{D}}$, except that all dynamic dependence formulas $D^n_X Y$ are taken as primitive now, rather than syntactical abbreviations. To interpret this language, we extend the earlier dynamic dependence models by associating each variable x with a topological space $\langle \mathfrak{D}_x, \tau_x \rangle$ on the set of values that x can take.

1 For details, see Baltag and van Benthem (2021a), which explores precise static analogues in an epistemic setting.

Definition 4.13: A tuple $\mathfrak{M} = \langle O, I, A, T, g\rangle$ is a *dynamic topological dependence model* (for short, *'dyn-topo-dep model'*) if

- O, I, A satisfy the same conditions as in dynamic dependence models.
- $T = \{\langle \mathfrak{D}_x, \tau_x\rangle \mid x \in \mathbb{V}\}$, where $\mathfrak{D}_x = \{o \in O \mid s(x) = o \text{ for some } s \in A\}$ and τ_x is a topology on $\mathfrak{D}_x$.
- $g : A \to A$ is a continuous function, in the sense of condition ($\star$) below.

To state the third condition more precisely, we need a further notion. Lifting from single variables to sets, given a dyn-topo-dep model $\mathfrak{M}$, we associate each non-empty set of variables X with a finite topological product space $\langle \mathfrak{D}_X, \tau_X\rangle$:

- $\mathfrak{D}_X := \{\langle s(x_i)\rangle_{x_i \in X} \mid s \in A\} \subseteq \prod_{x_i \in X} \mathfrak{D}_{x_i}$
- τ_X is the restriction to $\mathfrak{D}_X$ of the *product topology* on $\prod_{x_i \in X} \mathfrak{D}_{x_i}$, generated by the restriction to $\mathfrak{D}_X$ of all products $\prod_{x_i \in X} \tau_{x_i}$ of open sets.

In the extreme case of $X = \varnothing$, we get $\mathfrak{D}_X := \{\lambda\}$, where λ is the empty string, $\tau_X := \{\varnothing, \mathfrak{D}_X\}$ is the *discrete topology* and, for any assignment s, $s(\varnothing) := \lambda$.

Remark 4.2: Different orders of variables in X definitely give rise to different product topological spaces $\langle \mathfrak{D}_X, \tau_X\rangle$. In the reminder of the section, we always assume that there is a fixed order of variables, e.g., the one given by the enumeration $\mathbf{v}$ presented in Section 4.2.3. So, for simplicity, we write $\langle \mathfrak{D}_X, \tau_X\rangle$ for the unique topological spaces in harmony with the order. Also, with the order in mind, by slightly abuse of notations, for any sets of variables X and assignments s, we employ $s(X)$ for the *tuple of values* of variables in X, i.e., $s(X) := \langle s(x_i)\rangle_{x_i \in X} \in \mathfrak{D}_X$, which does *not* refer to the *set of values* of X.

Here is what we mean in Definition 4.13 by saying that the transition function g is 'continuous':

For any open U of the product space $\langle \mathfrak{D}_{\mathbb{V}}, \tau_{\mathbb{V}}\rangle$, the set of tuples of values
$\{s(\mathbb{V}) \mid g(s)(\mathbb{V}) \in U \text{ and } s \in A\}$ *is also an open in the space.* ($\star$)

The interior of a set $O \subseteq \mathfrak{D}_X$ in $\langle \mathfrak{D}_X, \tau_X\rangle$ is denoted by $\mathrm{Int}_X(O)$. Also, define $\tau(s(X)) := \{U \in \tau_X \mid s(X) \in U\}$, denoting the family of *open neighborhoods* of $s(X)$ in topology τ_X.[1]

Now we can introduce the semantics for the logic DCD.

1 The version presented here stays close to dynamic dependence models. Baltag and van Benthem (2021a) switch to models for DCD where variables are maps from abstract states to objects.

Definition 4.14: Let $\mathfrak{M} = \langle O, I, A, T, g\rangle$ be a dyn-topo-dep model. The *semantics for the language of* DCD is defined inductively by the following truth conditions. The clauses for $P\mathbf{x}$, $\neg, \wedge$ and $\bigcirc$ are the same as the earlier ones for DFD, and those for the modality $\mathsf{D}_X\varphi$ and the atoms $D^n_X Y$ are as follows:

$$s \vDash \mathsf{D}_X\varphi \quad \text{iff} \quad \exists U \in \tau(s(X))\ \forall t \in A\ (t(X) \in U \Rightarrow t \vDash \varphi)$$

$$\begin{aligned} s \vDash D^n_X Y \quad \text{iff} \quad & \exists U \in \tau(s(X))\ \forall t \in A : (t(X) \in U \Rightarrow \forall U' \in \tau(g^n(t)(Y)) \\ & \exists U'' \in \tau(t(X)) \forall t' \in A\ (t'(X) \in U'' \Rightarrow g^n(t')(Y) \in U')) \end{aligned}$$

The interpretation of $D^n_X Y$ expresses dynamic continuous dependence in its neighborhood version. Moreover, $\mathsf{D}_X\varphi$ states that the truth of φ is determined by some degree of approximation of the current values of X.

Here are a few useful notations. For any $\varphi \in \mathcal{L}_D$ and $X \subseteq \mathbb{V}$, $[\![\varphi]\!]_X := \{s(X) \in \mathbb{D}_X \mid s \in [\![\varphi]\!]\}$. It is easy to see that $[\![\varphi]\!]_{\mathbb{V}} = [\![\varphi]\!]$[1]. Also, by the truth condition for $\mathsf{D}_X\varphi$, we have that $[\![\mathsf{D}_X\varphi]\!] = \{t \in A \mid t(X) \in \mathrm{Int}_X([\![\varphi]\!]_X)\}$. In particular, when $X = \mathbb{V}$, $[\![\mathsf{D}_{\mathbb{V}}\varphi]\!] = \mathrm{Int}_{\mathbb{V}}[\![\varphi]\!]$. Mirroring standard topological semantics, $[\![\mathsf{D}_X\varphi]\!]$ is the interior of $[\![\varphi]\!]_X$ in the relevant topological space.

Important special case: Alexandroff models. *Alexandroff spaces* have topologies that are closed under arbitrary intersections, or equivalently, every point has a smallest open neighborhood. A dyn-topo-dep model $\mathfrak{M}$ is an *Alexandroff model* if for each $x \in \mathbb{V}$, $\langle \mathbb{D}_x, \tau_x\rangle$ is Alexandroff. Since the product of finitely many Alexandroff spaces is still Alexandroff (Arenas, 1999), all topological spaces $\langle \mathbb{D}_X, \tau_X\rangle$ used in our semantics on Alexandroff models are Alexandroff.

It is well-known that the τ_X in an Alexandroff space $\langle \mathfrak{D}_X, \tau_X\rangle$ coincides with the *upset topology* w.r.t. its *specialization preorder* $\leqslant_X \subseteq \mathfrak{D}_X \times \mathfrak{D}_X$:

For any $c, d \in \mathfrak{D}_X$, $c \leqslant_X d$ iff for all $U \in \tau_X$, if $c \in U$ then $d \in U$.

Accordingly, we obtain an equivalent relational semantics for our language.

Proposition 4.12: On Alexandroff models $\mathfrak{M}$, the truth conditions for $\mathsf{D}_X\varphi$ and $D^n_X Y$ in Definition 4.14 are equivalent to the following:

$$s \vDash \mathsf{D}_X\varphi \quad \text{iff} \quad \text{for all } t \in A,\ s \leqslant_X t \text{ implies } t \vDash \varphi$$

1 More precisely, $[\![\varphi]\!]_{\mathbb{V}} \subseteq O^{|\mathbb{V}|}$ is the class of tuples of values of all variables $\mathbb{V}$ given by admissible assignments satisfying φ, while $[\![\varphi]\!]$ is the class of admissible assignments satisfying φ. In the rest of the article, we will ignore this small difference.

$$s \vDash D^n_X Y \quad \text{iff} \quad \text{for all } t_1, t_2 \in A,\ s \leqslant_X t_1 \leqslant_X t_2 \text{ implies } g^n(t_1) \leqslant_Y g^n(t_2)$$

Proof The equivalence with our earlier semantics turns on the fact that each point in an Alexandroff space has a smallest open neighborhood. For then, truth throughout some open X-neighborhood of s amounts to truth in all points t with $s \leqslant_X t$, and the earlier clauses transcribe into the relational ones. ∎

4.5.1 Axiomatizing the logic of continuous dynamic dependence

We now proceed to axiomatizing the logic DCD by modifying our earlier proof calculus **DFD** of Table 4.1. For a start, this system is too strong for validity on dyn-topo-dep models. For instance, the topological operators D_X are essentially **S4**-modalities. The modal **S5**-principle $\neg\mathsf{D}_X\varphi \to \mathsf{D}_X\neg\mathsf{D}_X\varphi$ (D-5) only holds on special topologies, though we do have this principle in general when $X = \varnothing$, as $\mathsf{D}_\varnothing$ is the universal modality. Another principle that fails is D-Introduction$_1$: when a fact is true at some point, it need not be true in an open neighborhood around that point. Finally, since formulas $D^n_X Y$ are primitive now, we also need modify the principles for dependence atoms in the system **DFD**.

Despite these differences, the surprising fact is that the bulk of the system **DFD** represents valid insights under the topological reading. For instance, the axiom Determinism $D^1_{\mathbb{V}} x$, and in particular, its consequence $D^1_{\mathbb{V}}\mathbb{V}$, captures the continuity of the dynamic transition function g in a natural way.[1] Finally, is there also a principle characterizing the continuity aspect of dynamic dependence? Indeed, the calculus **DFD** does contain such an axiom, viz. Transfer:

$$D^n_X Y \wedge \bigcirc^n \mathsf{D}_Y \varphi \to \mathsf{D}_X \bigcirc^n \varphi$$

which is still valid on dyn-topo-dep models because of continuity considerations.

Table 4.2 displays a proof system **DCD** for the logic DCD. It is important to notice here that, although many principles are syntactically the same as those in Table 4.1, they have quite different meanings in the topological setting. A few more details of this meaning shift will follow after we have stated the next result.

Proposition 4.13: The proof system **DCD** is sound w.r.t. dyn-topo-dep models.

1 It may be worth mentioning here that the topologically valid principles $D^1_{\mathbb{V}}\mathbb{V}$ and Dyn-Trans of **DFD** derive the implication $\bigcirc\mathsf{D}_{\mathbb{V}}\varphi \to \mathsf{D}_{\mathbb{V}}\bigcirc\varphi$, which is a direct analogue of the key continuity axiom of Dynamic Topological Logic (Kremer and Mints, 2007).

Table 4.2 The Proof System **DCD**.

I	**Axioms and Rules of Classical Propositional Logic**
II	**Axioms and Rules for $\bigcirc$:** Part II of Table 4.1
III	**Axioms and Rules for D**
D-Distribution	$\mathsf{D}_X(\varphi \to \psi) \to (\mathsf{D}_X\varphi \to \mathsf{D}_X\psi)$
D-Introduction	$D^n_X Y \to \mathsf{D}_X D^n_X Y$
D-T	$\mathsf{D}_X\varphi \to \varphi$
D-4	$\mathsf{D}_X\varphi \to \mathsf{D}_X\mathsf{D}_X\varphi$
D-w5	$\neg\mathsf{D}_\varnothing\varphi \to \mathsf{D}_\varnothing\neg\mathsf{D}_\varnothing\varphi$
D-Necessitation	From φ, infer $\mathsf{D}_X\varphi$
IV	**Axioms for $D^n_X Y$**
Dep-Inclusion	$D_X Y$, for all $Y \subseteq X$
Dep-Additivity	$D^n_X Y \wedge D^n_X Z \to D^n_X(Y \cup Z)$
Dyn-Trans	$D^n_X Y \wedge \bigcirc^n D^m_Y Z \to D^{n+m}_X Z$
Cont-Determinism	$D^1_{\mathbb{V}}\mathbb{V}$
V	**Interaction Axioms:** Part V of Table 4.1

We will merely highlight the key reasons for the validity of the principles.

Proof Axioms D-T and D-4 hold by the fact that all D_X are **S4**-modalities. Also, as $\mathsf{D}_\varnothing$ is the universal modality, we have D-w5 and D-$\bigcirc$ (Group V) as well. D-Introduction holds by the fact that $D^n_X Y$ is true globally on some open of a point when it is true at that point by its truth condition in Definition 4.14.

The validity of all other non-trivial principles stems from the continuity of the transition function g and known properties of product topologies (cf. Armstrong, 1983):

- Cont-Determinism follows directly from the continuity of the function g.
- Dep-Inclusion holds by the fact that *projections* w.r.t. the product topology are continuous.
- Dep-Additivity follows from the *universality property* for product topology.
- Both Dyn-Trans and Transfer follow from the fact that the composition of continuous functions is continuous as well. ∎

4.5.2 Completeness of **DCD**: introducing dynamic preorder models

Our main remaining task is to verify that the calculus **DCD** is complete w.r.t. the class of dyn-topo-dep models. In what follows, Alexandroff topologies play a central role: completeness holds immediately if we can show that **DCD** is complete w.r.t. Alexandroff models. To show the latter, our strategy is similar to the completeness proof for **DFD**, but the route is a bit different, and simpler.

First, we introduce modal-style *dynamic preorder models* (for short, *preorder models*), an analogue of the general relational models in Definition 4.10.

Definition 4.15: A *preorder model* is a tuple $\mathcal{M} = \langle W, \leqslant_X, g, D^n_X Y, P\mathbf{x}\rangle$,[1] where

- W is a non-empty set of possible worlds.
- For each set $X \subseteq \mathbb{V}$, $\leqslant_X \subseteq W \times W$ is a binary relation.
- $g : W \to W$ is a total function.
- For each n-ary predicate symbol P and each world $s \in W$, P^s is an n-ary relation on objects.
- For each $s \in W$, $(D^n_X Y)^s \subseteq \mathcal{P}(\mathbb{V}) \times \mathcal{P}(\mathbb{V})$ is a relation between sets of variables X and Y.

These components are required to satisfy the following conditions:

P1 All $\leqslant_X$ are preorders, i.e., reflexive and transitive relations.

P2 $\leqslant_\varnothing$ is the universal relation.

P3 All relations $(D^n_X Y)^s$ satisfy the above principles of Dep-Inclusion, Dep-Additivity, Dyn-Transitivity and Cont-Determinism.

P4 For all $s, t \in W$, if $s \leqslant_X t$ and $(D^n_X Y)^s$, then $g^n(s) \leqslant_Y g^n(t)$ and $(D^n_X Y)^t$.

It is useful to note the following consequences of properties P3 and P4:

- If $Y \subseteq X$, then $\leqslant_X \subseteq \leqslant_Y$, i.e., $s \leqslant_X t$ implies $s \leqslant_Y t$.
- The function g preserves $\leqslant_\mathbb{V}$, i.e., $s \leqslant_\mathbb{V} t$ implies $g(s) \leqslant_\mathbb{V} g(t)$.[2]

The semantics for the logic DCD on preorder models is as follows.

Definition 4.16: Let $\mathcal{M} = \langle W, \leqslant_X, g, D^n_X Y, P\mathbf{x}\rangle$ be a preorder model, and let $s \in W$. The truth conditions for atoms $P\mathbf{x}$, $D^n_X Y$, Boolean connectives and $\bigcirc$ are straightforward, and that for the modality D_X is standard:

1 Strictly speaking, the tuple defining the model contains families of relations and of atoms, but we will stick with this simpler notation.

2 As usual, monotonicity plays the role of continuity in a preorder setting.

$$\mathcal{M}, s \vDash \mathsf{D}_X\varphi \quad \text{iff} \quad \text{for each } t \in W,\ s \leqslant_X t \text{ implies } \mathcal{M}, t \vDash \varphi.$$

Given these truth conditions plus the fact that Definition 4.15 has just built in the validity of many axioms, it is not hard to check the following:

Proposition 4.14: The system **DCD** is sound w.r.t. dynamic preorder models.

4.5.3 The canonical model for DCD

To show that the calculus **DCD** is complete w.r.t. dynamic preorder models, a standard modal argument suffices, similar to the one for **DFD** in Section 4.3.3.

Definition 4.17: The *canonical preorder model* of **DCD** is the tuple $\mathcal{M}^c = \langle W^c, g^c, \leqslant^c_X, V^c\rangle$ satisfying the following four conditions:

- W^c is the set of all maximal **DCD**-consistent sets.
- For all $s \in W^c$, $g^c(s) = \{\varphi \in \mathcal{L}_D \mid \bigcirc\varphi \in s\}$.
- For all $s, t \in W^c$ and $X \subseteq \mathbb{V}$, $s \leqslant^c_X t$ iff $\mathsf{D}_X s \subseteq t$.
- $s \in V^c(D^n_X Y)$ iff $D^n_X Y \in s$, and $s \in V^c(P\mathbf{x})$ iff $P\mathbf{x} \in s$.

Just as with **DFD**, we will focus on the *generated canonical submodel* w.r.t. the relation $\leqslant^c_\varnothing$. Also just as before, it can be shown that g^c is well-defined:

Proposition 4.15: In the canonical preorder model $\mathcal{M}^c$, $g^c(s) \in W^c$.

Now, reasoning as in the proof of Proposition 4.11, we can verify the following

Proposition 4.16: The canonical preorder model is a dynamic preorder model.

Next, an Existence Lemma and Truth Lemma can be shown in modal style:

Lemma 4.5: Let $\mathcal{M}^c$ be the canonical preorder model, and $s \in W^c$. We have:

$$\text{if } \widehat{\mathsf{D}}_X\varphi \in s, \text{ then there exists } t \in W^c \text{ such that } s \leqslant^c_X t \text{ and } \varphi \in t.$$

Lemma 4.6: Let $\mathcal{M}^c$ be the canonical preorder model, and $s \in W^c$. Then:

$$\mathcal{M}^c, s \vDash \varphi \text{ iff } \varphi \in s.$$

Together, these observations establish the following result.

Theorem 4.4: The system **DCD** is complete w.r.t. dynamic preorder models.

This part of the completeness proof was routine: now, the real work starts.

4.5.4 Equivalence of Alexandroff models and preorder models

In this part, we aim to clarify the relation between the class of Alexandroff models and dynamic preorder models. In particular, as we will see,

the logic of Alexandroff models is exactly the logic of preorder models,

from which completeness of **DCD** w.r.t. dyn-topo-dep models follows. The displayed fact follows from two model constructions that we will consider separately.

First, we show how each Alexandroff model induces a preorder model.

Definition 4.18: For an Alexandroff model $\mathfrak{M} = \langle O, I, A, T, g\rangle$, the *induced dynamic preorder model* $\mathcal{M}^{\mathfrak{M}} = \langle W^{\mathfrak{M}}, \leqslant_X^{\mathfrak{M}}, g^{\mathfrak{M}}, D_X^n Y, P\mathbf{x}\rangle$ is given as follows:

- $W^{\mathfrak{M}} := A$, and $g^{\mathfrak{M}} := g$.
- For each $X \subseteq \mathbb{V}$, the relation $\leqslant_X^{\mathfrak{M}}$ is the specialization preorder of the topological space $\langle \mathfrak{D}_X, \tau_X\rangle$.
- Truth values of atoms $P\mathbf{x}$ are given by the valuation of the Alexandroff model.
- $(D_X^n Y)^s$ iff for all $t_1, t_2 \in A$, $s \leqslant_X^{\mathfrak{M}} t_1 \leqslant_X^{\mathfrak{M}} t_2$ implies $g^n(t_1) \leqslant_Y^{\mathfrak{M}} g^n(t_2)$.

The last clause of this definition is essentially the relational-style truth condition stated in Proposition 4.12 for $D_X^n Y$ in Alexandroff models.

To see that the induced model is a preorder model, it is crucial to note that $D_{\mathbb{V}}^1 \mathbb{V}$ holds globally.[1] By induction on $\varphi \in \mathcal{L}_D$, the following can then be shown.

Proposition 4.17: For each Alexandroff model $\mathfrak{M}$ and $\varphi \in \mathcal{L}_D$,

$$\mathfrak{M}, s \vDash \varphi \text{ iff } \mathcal{M}^{\mathfrak{M}}, s \vDash \varphi.$$

For the other direction, we have a less simple result requiring more care.

Theorem 4.5: Each preorder model is a *p*-morphic image of the preorder model induced by some Alexandroff model.

Proof The proof strategy is similar to the unraveling found with Theorem 4.2, but details of the route are different in the topological setting.

Let $\mathcal{M} = \langle W, \leqslant_X, g, D_X^n Y, P\mathbf{x}\rangle$ be an arbitrary preorder model. We will construct a new special 'tree-like' preorder model

1 Suppose otherwise. Then there are $s, t_1, t_2 \in W^{\mathfrak{M}}$ with $s \leqslant_{\mathbb{V}}^{\mathfrak{M}} t_1 \leqslant_{\mathbb{V}}^{\mathfrak{M}} t_2$ and $g(t_1) \not\leqslant_{\mathbb{V}}^{\mathbb{M}} g(t_2)$: i.e., there is an open $U \in \tau_{\mathbb{V}}$ with $g(t_1) \in U$ and $g(t_2) \notin U$. As g is continuous, $g^{-1}[U]$ is also an open, which contains t_1 but not t_2. So, $t_1 \not\leqslant_{\mathbb{V}}^{\mathfrak{M}} t_2$, a contradiction.

$$\mathcal{M}' = \langle \mathcal{H}, \leqslant'_X, g', D^n_X Y, P\mathbf{x} \rangle$$

that can be associated with a concrete topological Alexandroff model, and which is also designed so that the given $\mathcal{M}$ is a p-morphism image of $\mathcal{M}'$.

The set $\mathcal{H}$ of *states* consists of all finite *histories* $h = \langle s_0, \bigtriangleup_1, s_1, \cdots, \bigtriangleup_n, s_n \rangle$ (for arbitrary lengths $n \in \mathbb{N}$) such that:

- For each $i \leqslant n$, $s_i \in W$, and $\bigtriangleup_i$ is $X_i \subseteq \mathbb{V}$ or g.
- (a) If $\bigtriangleup_i$ is X_k, then $s_{k-1} \leqslant_{X_k} s_k$; (b) if $\bigtriangleup_k$ is g, then $g(s_{k-1}) = s_k$.

By $last(h)$ we denote the last state in the history h. The *dynamic transition function* g' on $\mathcal{H}$ is now given by the following:

- $g'(h) = \langle h, g, last(h') \rangle$.

Next, in order to present $\mathcal{M}'$ completely, it remains to define suitable relations $\leqslant'_X$ as well as a valuation for atoms $P\mathbf{x}$ and $D^n_X Y$. Having done that, we will show how there is an Alexandroff space inducing this structure.

First, we supply *ranges of values* for variables. For each variable $x \in \mathfrak{V}$, set

$$\mathfrak{D}_x := \{(h, x) \mid h \in \mathcal{H}\}$$

As in Definition 4.18, this allows us to also view histories as assignments by setting

$$h(x) := (h, x)$$

The *interpretation map* I is as follows, for each n-ary predicate symbol P:

- $I(P) := \{\langle (h, x_1), \cdots, (h, x_n) \rangle \mid h \in \mathcal{H}$ *with* $\mathcal{M}, last(h) \vDash Px_1 \cdots x_n\}$.

Next, we move to the relations $\leqslant'_X$. First, for each variable $x \in \mathbb{V}$, we define a one-step relation $hR_x h'$ between histories as follows:

- h is of the form $\langle h_0, g, t_1, g, t_2, \cdots, g, t_n \rangle$ and h' is $\langle h'_0, g, s_1, g, s_2, \cdots, g, s_n \rangle$ such that $h'_0 = \langle h_0, Y, last(h'_0) \rangle$ for some $Y \subseteq \mathbb{V}$ with $last(h_0) \vDash D^n_Y x$.

Let R^*_x be the reflexive-transitive closure of R_x. We define $\leqslant'_x$ on $\mathfrak{D}_x$ as follows:

$$(h, x) \leqslant'_x (h', x) \text{ iff } hR^*_x h'.$$

Finally, the *interpretation of dependence atoms* $D^n_X Y$ in the model $\mathcal{M}'$ arises exactly in the way described in Definition 4.18.[1]

Now we are in a position to supply an associated Alexandroff model. Note that the relations $\leqslant'_x$ are clearly reflexive and transitive because R^*_x is. Moreover, it is simple to

1 Crucially, then, its interpretation is not as free as in preorder models in general.

see that the relation is *anti-symmetric* as well: moving along the relation h_x, sizes of histories (i.e., the number of elements in histories) become larger and larger, so $h \leqslant'_x h'$ and $h' \leqslant'_x h$ are followed by $h = h'$. Therefore, the relational frame $\langle \mathfrak{D}_x, \leqslant'_x \rangle$ induces an Alexandroff T_0-space $\langle \mathfrak{D}_x, \tau_x \rangle$[1] whose specialization preorder is exactly the relation $\leqslant'_x$ itself.

With the preceding observation, for the rest of the completeness proof, we can concentrate on the structure $\mathcal{M}' = \langle \mathcal{H}, g', \leqslant'_X, D^n_X y, P\mathbf{x} \rangle$ defined here.

Lemma 4.7: The tuple $\mathcal{M}' = \langle \mathcal{H}, g', \leqslant'_X, D^n_X y, P\mathbf{x} \rangle$ is a dynamic preorder model.

Proof We must check all the conditions in Definition 4.15.

Claim 1. All relations $\leqslant'_X$ are preorders, and $\leqslant'_\emptyset$ is the universal relation.

Proof By construction, it holds obviously that $\leqslant'_\emptyset$ is the universal relation. Let $X \subseteq \mathbb{V}$ be an arbitrary set of variables. Recall that all those $\langle \mathfrak{D}_x, \tau_x \rangle$ are essentially Alexandroff topological T_0-spaces. As noted in (Mahdi, 2010), the specialization order of a product topology of finite many Alexandroff topological T_0-spaces is exactly the intersection of the original specialization orders. Immediately, w.r.t. the specialization order $\leqslant'_X$ of the resulting product topological space $\langle \mathfrak{D}_X, \tau_X \rangle$, it holds that:

$$(h, X) \leqslant'_X (h', X) \text{ iff } (h, x) \leqslant'_x (h', x) \text{ for all } x \in X.$$

Since all $\leqslant'_x$ are preorders, it is easy to see that $\leqslant'_X$ is a preorder as well. ∎

Claim 2. If $h \leqslant'_X h'$ and $last(h) \vDash D^n_X Y$, then $g'^n(h) \leqslant'_Y g'^n(h')$ and $last(h') \vDash D^n_X Y$.

Proof The case that $h = h'$ is trivial. For the more general situations, it suffice to analyze a single step. Consider $hR_x h'$ for all $x \in X$. We now show $g'^n(h) \leqslant'_Y g'^n(h')$ and $last(h') \vDash D^n_X Y$. Notice that $h'_0 = \langle h_0, Z, last(h'_0) \rangle$ for some $Z \subseteq \mathfrak{V}$ and $last(h_0) \vDash D^i_Z X$. Also, $last(h) \vDash D^n_X Y$ indicates that $last(h_0) \vDash \bigcirc^i D^n_X Y$. Therefore, by P3 of Definition 4.15, $last(h_0) \vDash D^{i+n}_Z Y$. So, $g'^n(h) \leqslant'_Y g'^n(h')$. Moreover, as $last(h_0) \leqslant_Z last(h'_0)$, from P4 in Definition 4.15 and $last(h_0) \vDash D^i_Z X$ it follows $last(h) \leqslant_X last(h')$. Now, using P4 again, from $last(h) \vDash D^n_X Y$ we have $last(h') \vDash D^n_X Y$. ∎

Claim 3. $(D^n_X Y)^h$ satisfy the conditions 'Dep-Inclusion', 'Dep-Additivity', 'Dyn-Trans' and 'Cont-Determinism'.

1 In a T_0-space, any two different points in can be separated by an open set.

Proof By Definition 4.18, $(D^n_X Y)^h$ iff $h \leqslant'_X h' \leqslant'_X h''$ entails $g'^n(h') \leqslant'_Y g'^n(h'')$. Also, as $\mathcal{M}$ is a preorder model, all $(D^n_X Y)^s$ satisfy these conditions. So, it suffices to show that following assertions are equivalent for all histories h:

(a) For all histories $h', h'' \in \mathcal{H}$, $h \leqslant'_X h' \leqslant'_X h'' \Rightarrow g'^n(h') \leqslant'_Y g'^n(h'')$

(b) $\mathcal{M}, last(h) \vDash D^n_X Y$

From (a) to (b). Let $last(h) := s$ and $h' := \langle h, X, s\rangle$. We are going to prove $\mathcal{M}, s \vDash D^n_X y$. As $\leqslant'_X$ is reflexive, $h \leqslant'_X h \leqslant'_X h'$. So, $g'^n(h) \leqslant'_Y g'^n(h')$. Similar to that of Theorem 4.2, it holds that $g'^n(h) R_y g'^n(h')$ for all $y \in Y$. So, by the definition of R_y, we have $\mathcal{M}, last(h) \vDash D^n_X y$ for all $y \in Y$. Consequently, from property Dep-Additivity, we have $\mathcal{M}, last(h) \vDash D^n_X Y$.

From (b) to (a). For the other direction, suppose that $\mathcal{M}, last(h) \vDash D^n_X Y$. Let h', h'' be two histories s.t. $h \leqslant'_X h' \leqslant'_X h''$. Using Claim 2 twice, it is easy to see that $g'^n(h') \leqslant'_Y g'^n(h'')$. ∎

This completes the proof for Lemma 4.7. ∎

Lemma 4.8: The function $last : \mathcal{H} \to W$ mapping a history to its last state is a surjective p-morphism from $\mathcal{M}'$ to $\mathcal{M}$.

Proof The *surjectivity* of function $last$ is straightforward: for any $s \in W$, we have $h = \langle s\rangle \in \mathcal{H}$. Next, by similar reasoning to that of Lemma 4.4, we can show the following *forth and back conditions* for $\leqslant'_X$ and g':

- If $h \leqslant'_X h'$, then $last(h) \leqslant_X last(h')$.
- If $last(h) \leqslant_X s$, then there is a history h' such that $h \leqslant'_X h'$ and $last(h') = s$.
- If $g'(h) = h'$, then $g(last(h)) = last(h')$.
- If $g(last(h)) = s$, then there is a history h' with $g'(h) = h'$ and $last(h') = s$.

Finally, it remains to consider the valuation function. The case for $Px_1 \cdots x_n$ is simple, which holds directly by the definition of interpretation map I: $\mathcal{M}', h \vDash Px_1 \cdots x_n$ iff $\langle\langle h, x_1\rangle, \cdots, \langle h, x_n\rangle\rangle \in I(P)$ iff $\mathcal{M}, last(h) \vDash Px_1 \cdots x_n$. Moreover, the case for $D^n_X y$ essentially has already been proved in that of Claim 3.

This completes the proof for Lemma 4.8. ∎

Now the proof of Theorem 4.5 is complete. ∎

Combining Proposition 4.17 and Theorem 4.5, it follows immediately that:

Proposition 4.18: The same $\mathcal{L}_D$-formulas are valid on arbitrary dynamic preorder models and on Alexandroff preorder models.

Therefore, using also Proposition 4.13 and Theorem 4.4, we obtain:

Theorem 4.6: The system **DCD** is sound and complete for Alexandroff models.

Finally, as **DCD** was sound for all dyn-topo-dep models, it follows that:

Theorem 4.7: The system **DCD** is complete w.r.t. dyn-topo-dep models.

4.6 Summary and future work

Summary. This chapter has developed two logical systems for modeling and reasoning about dependence in dynamical systems. Our original motivation came from patterns of social interactions in game-like situations, but our systems can also be viewed as more general logics of dependence. In particular, our proposals took both the dependence quantifiers $\mathsf{D}_X\varphi$ and the dynamic dependence formulas $D^n_X y$ into account:

- Dependence quantifiers $D_X\varphi$ together with the temporal operator $\bigcirc$ captured dependence across successive actions.
- Dynamic dependence formulas characterized dependencies between actions of agents in social scenarios.

The two sorts of dependence also had various interactions captured by our logics which could then in principle be used to analyze the graph games of our earlier chapters, though we did not pursue this direction here.

Our first system DFD was concerned with dependence between variables and step-by-step temporal progression. Axiomatizing this logic required the use of abstract modal methods, and the proof patterns presented here may well have a much larger scope. As an aside, we also presented some results on first-order translation and bisimulation that clarify the expressive power of DFD. Also, we showed that its fragment not containing $\mathsf{D}_\varnothing\varphi$ or $D^n_\varnothing y$ is decidable.

Next, adding the topological structure found in many dynamical systems, we presented the system DCD with a language enriched with modalities for topological interior that can also reason about continuous transition functions and continuous forms of dynamic dependence. This language has a proof system with new topological content, and we showed its completeness making adjustments to the DFD case in order accommodate the topological structure.

In this way, we have connected several varieties of modal logic: topological logics, modal dependence logic, and temporal logic, or stated differently, we have linked up between modal dependence logic and dynamic temporal logic.

Further directions. Several open problems remain about the two systems presented here. A striking one is the decidability and computational complexity of the whole logics. Another obvious issue is a proof-theoretic Gentzen-style treatment of the two systems, as has already been given for basic modal dependence logic.

Next, our languages are relatively poor. One natural direction would be adding the standard *temporal future operator* which allows us to reason about the eventual long-term behavior of dynamical systems. This may well be a challenging direction, given the experience in dynamic topological logic (Kremer and Mints, 2007). Another natural language extension would move from the bare variables that we have employed to allow *function terms* of various sorts: static and dynamic.

Also, our models are rather simple, in that they identify states with variable assignments. This feature can be generalized in various ways, and one setting where more abstract states naturally surface is in exploring the rich topological structure that we have introduced. We have studied one particular form of continuous dynamic dependence, but many alternatives make sense in empirical science, such as the 'stable' and 'sensitive' readings of dynamic continuous dependence introduced in Section 4.4.1.

Finally, dynamical systems remain highly abstract in this theoretical and foundational chapter. Much more structure will come to light when we apply the logics presented here to concrete settings such as dynamic Markov models for social processes over time.

Appendix: Decidability of a fragment of DFD

We will show that the fragment of logic DFD without $D^n_{\varnothing} y$ and $\mathsf{D}_{\varnothing}\varphi$ has the finite model property, making its decidable. We begin with the following syntactic notion:

Definition 4.19: Let φ be an $\mathcal{L}_{\mathsf{D}}$-formula. The *temporal depth* of φ, $\mathbf{td}(\varphi)$, is recursively defined as follows:

- $\mathbf{td}(Px_1 \cdots x_n) = 0$
- $\mathbf{td}(\neg\varphi) = \mathbf{td}(\mathsf{D}_X\varphi) = \mathbf{td}(\varphi)$
- $\mathbf{td}(\varphi \wedge \psi) = \max\{\mathbf{td}(\varphi), \mathbf{td}(\psi)\}$
- $\mathbf{td}(\bigcirc\varphi) = \mathbf{td}(\varphi) + 1$
- $\mathbf{td}(D^n_X y) = n$

Also, for a set $\Phi \subseteq \mathcal{L}_{\mathsf{D}}$ of formulas, $\mathbf{td}(\Phi) = \max\{\mathbf{td}(\varphi) \mid \varphi \in \Phi\}$. Moreover, let us introduce another notion of '*strong closure*' for finite sets of $\mathcal{L}_{\mathsf{D}}$-formulas:

Definition 4.20: Let Φ be a finite set of $\mathcal{L}_{\mathsf{D}}$-formulas with $\mathbf{td}(\Phi) = k$. We denote by $\mathbb{V}_{\Phi}$ the set of all variables occurring in Φ. The set Φ is ***strongly closed*** if:

($SC1$). If $\bigcirc^{n+1}\varphi \in \Phi$, then $\bigcirc^n\varphi \in \Phi$.

($SC2$). If $\bigcirc^n\mathsf{D}_X\varphi \in \Phi$, then $\bigcirc^n\varphi \in \Phi$.

($SC3$). If $\bigcirc^n\neg\varphi \in \Phi$, then $\bigcirc^n\varphi \in \Phi$.

($SC4$). If $\bigcirc^n(\varphi_1 \wedge \varphi_2) \in \Phi$, then $\bigcirc^n\varphi_1 \in \Phi$ and $\bigcirc^n\varphi_2 \in \Phi$.

($SC5$). For all non-empty $X \subseteq \mathbb{V}_{\Phi}$ and $y \in \mathbb{V}_{\Phi}$, $D^k_X y \in \Phi$.

($SC6$). If $\bigcirc^n D^m_X y \in \Phi$, then $\bigcirc^{n'} D^{m'}_{X'} y' \in \Phi$ for all $n' + m' = n + m$, $\varnothing \neq X' \subseteq \mathbb{V}_{\Phi}$ and $y' \in \mathbb{V}_{\Phi}$.

($SC7$). If $\varphi \in \Phi$ includes no operator D_X for any $X \subseteq \mathbb{V}_{\Phi}$, then $\mathsf{D}_X\varphi \in \Phi$ for all non-empty $X \subseteq \mathbb{V}_{\Phi}$.

($SC8$). If $\bigcirc^m\mathsf{D}_X \bigcirc^n \varphi \in \Phi$, then $\bigcirc^{m'}\mathsf{D}_Y \bigcirc^{n'} \varphi \in \Phi$ for all non-empty $Y \subseteq \mathbb{V}_{\Phi}$ and $m' + n' = m + n$.

For a finite set Ψ of formulas, $Cl(\Psi)$ denotes its ***strong closure***, i.e., the smallest closed set of formulas containing Ψ. When Ψ is a singleton $\{\varphi\}$, we use $Cl(\varphi)$ for $Cl(\{\varphi\})$.

It is not hard to see that the strong closure of a finite set is also finite. Also, by a simple induction on the structure of formulas, it holds that:

Proposition 4.19: For any $\Psi \subseteq \mathcal{L}_{\mathsf{D}}$, if it satisfies conditions ($SC1$)-($SC4$), then it is closed under subformulas.

Another simple observation is as follows:

Proposition 4.20: For any $\Psi \subseteq \mathcal{L}_{\mathsf{D}}$ satisfying $(SC6)$-$(SC8)$, if $\bigcirc^n D^m_X y \in \Psi$, then $\bigcirc^{n'} \mathsf{D}_{X'} D^{m'}_{X'} y' \in \Psi$ for all $n' + m' = n + m$, non-empty $X' \subseteq \mathbb{V}_\Psi$ and $y' \in \mathbb{V}_\Psi$.

Additionally, it also holds that:

Proposition 4.21: For any $\Psi \subseteq \mathcal{L}_{\mathsf{D}}$ satisfying $(SC1)$, $(SC5)$ and $(SC6)$, if $\mathbf{td}(\Psi) = k$, then for all $j \leqslant k$, and non-empty $X \subseteq \mathbb{V}_\Psi$ and $y \in \mathbb{V}_\Psi$, it holds that $D^j_X y \in \Psi$.

We now introduce some useful notions. Let $\varphi \in \mathcal{L}_{\mathsf{D}}$. We call $\langle +, \varphi \rangle$ and $\langle -, \varphi \rangle$ *signed formulas*. For brevity, we often write $+\varphi$ for $\langle +, \varphi \rangle$, and $-\varphi$ for $\langle -, \varphi \rangle$. A *pseudo-atom* is a set of signed formulas. For any pseudo-atom α, define $|\alpha| := \{\varphi \mid +\varphi \in \alpha \text{ or } -\varphi \in \alpha\}$, and say that α *is strong closed* if $|\alpha|$ is strong closed. Also, set $\mathbf{td}(\alpha) := \mathbf{td}(|\alpha|)$. Now we define a notion of '*identification*' for pseudo-atoms α (written $\mathbf{I}(\alpha)$):

- When $\alpha = \varnothing$, $\mathbf{I}(\alpha) := D_x x$; and
- When $\alpha \neq \varnothing$, $\mathbf{I}(\alpha) := \bigwedge_{+\varphi_i \in \alpha} \varphi_i \wedge \bigwedge_{-\varphi_j \in \alpha} \neg \varphi_j$.

For instance, when $\alpha = \{-\varphi_1, +\varphi_2, -\varphi_3\}$, $\mathbf{I}(\alpha) = \neg\varphi_1 \wedge \varphi_2 \wedge \neg\varphi_3$. Furthermore, we say a pseudo-atom α is *consistent* just in case its identification $\mathbf{I}(\alpha)$ is consistent.

Definition 4.21: Given a strong closed $\Phi \subseteq \mathcal{L}_{\mathsf{D}}$, a Φ*-atom* is $\varnothing$ or any consistent strong closed pseudo-atom α with $|\alpha| \subseteq \Phi$ and $\mathbb{V}_{|\alpha|} = \mathbb{V}_\Phi$.

For any strong closed $\Phi \subseteq \mathcal{L}_{\mathsf{D}}$ and non-empty Φ-atom α, condition $\mathbb{V}_{|\alpha|} = \mathbb{V}_\Phi$ together with the consistency of α ensures $+D_X x \in \alpha$ for all $x \in X \subseteq \mathbb{V}_\Phi$.

In what follows, we also write $+D^n_X Y \in \alpha$ when $+D^n_X y \in \alpha$ for all $y \in Y$. Now, let us introduce the following notion of 'Φ*-model*' for strong closed $\Phi \subseteq \mathcal{L}_{\mathsf{D}}$:

Definition 4.22: Let $\Phi \subseteq \mathcal{L}_{\mathsf{D}}$ be a strong closed set. The Φ*-model* is a tuple $\mathcal{M}^\Phi = \{W^\Phi, g^\Phi, =^\Phi_X, V^\Phi\}$ defined as follows:

M1. $W^\Phi = \{\alpha \mid \alpha \text{ is a } \Phi\text{-atom}\}$.

M2. $g^\Phi(\alpha) = \{+\varphi \mid + \bigcirc \varphi \in \alpha\} \cup \{-\varphi \mid - \bigcirc \varphi \in \alpha\}$.

M3. For all non-empty $X \subseteq \mathbb{V}_\Phi$, $\alpha =^\Phi_X \beta$ iff

- M3.1. For all $D^n_X Y$, $+D^n_X Y \in \alpha \Leftrightarrow +D^n_X Y \in \beta$, and
- M3.2. When $+D^n_X Y \in \alpha$ (or equivalently, $+D^n_X Y \in \beta$), for all $\bigcirc^n \mathsf{D}_Y \varphi \in \mathcal{L}_{\mathsf{D}}$, it holds $+ \bigcirc^n \mathsf{D}_Y \varphi \in \alpha \Leftrightarrow + \bigcirc^n \mathsf{D}_Y \varphi \in \beta$.

M4. For all $P\mathbf{x}$, $V^\Phi(P\mathbf{x}) = \{\alpha \mid +P\mathbf{x} \in \alpha\}$. For all dynamic dependence formulas $D^n_X y$, $V^\Phi(D_X y)$ is defined as follows:

M4.1. For all $n \in \mathbb{N}$, $V^{\Phi}(D^n_{\mathbb{V}_\Phi} x) = W^{\Phi}$

M4.2. For all $y \in X \subseteq \mathbb{V}_\Phi$, $V^{\Phi}(D_X y) = W^{\Phi}$

M4.3. For all other $D^n_X y$, $V^{\Phi}(D^n_X y)$ is the smallest set satisfying the following:

M4.3.1. If $+D^n_X y \in \alpha$, then $\alpha \in V^{\Phi}(D^n_X y)$

M4.3.2. If $+D^m_X \mathbb{V}_\Phi \in \alpha$ for some $m \leqslant n$, then $\alpha \in V^{\Phi}(D^n_X y)$.

Proposition 4.22: The valuation function in the definition above is well-defined.

Here are some observations and comments on Φ-models:

- Model $\mathcal{M}^{\Phi}$ is finite whenever Φ is so.
- For all $X \neq \varnothing$, if $\alpha =^{\Phi}_X \beta$, then $\alpha = \varnothing$ iff $\beta = \varnothing$.
- Clauses M4.3.1 and M4.3.2 ensure that $\alpha \in V^{\Phi}(D^n_X y)$ for all valid $D^n_X y$.
- For all $D^n_X y$ with $n \leqslant \mathbf{td}(\alpha)$, it holds $\alpha \in V^{\Phi}(D^n_X y)$ iff $+D^n_X y \in \alpha$.

Here, it is crucial to point out that the function g^{Φ} in the Φ-model is well-defined on the domain W^{Φ}, in the sense of the following:

Proposition 4.23: For a strong closed set $\Phi \subseteq \mathcal{L}_{\mathsf{D}}$, if α is a Φ-atom, then $g^{\Phi}(\alpha)$ is also a Φ-atom. As a consequence, $g^{\Phi}(\alpha) \in W^{\Phi}$.

Proof Assume that α is a Φ-atom. The case that $g^{\Phi}(\alpha) = \varnothing$ is trivial. We now consider $g^{\Phi}(\alpha) \neq \varnothing$. Then, obviously $\mathbb{V}_\Phi = \mathbb{V}_{|\alpha|} = \mathbb{V}_{|g^{\Phi}(\alpha)|}$. Now it suffices to show that: (1). $g^{\Phi}(\alpha)$ is consistent; (2). $|g^{\Phi}(\alpha)| \subseteq \Phi$; and (3). $g^{\Phi}(\alpha)$ is strong closed. Let us begin.

(1). As α is consistent, $\bigwedge_{+\bigcirc\varphi_i \in \alpha} \bigcirc\varphi_i \wedge \bigwedge_{-\bigcirc\varphi_j \in \alpha} \neg\bigcirc\varphi_j$ is consistent as well. Therefore, $\bigwedge_{+\bigcirc\varphi_i \in \alpha} \varphi_i \wedge \bigwedge_{-\bigcirc\varphi_j \in \alpha} \neg\varphi_j$ is consistent. By the definition of g^{Φ}, it follows that $\mathbf{I}(g^{\Phi}(\alpha)) = \bigwedge_{+\varphi_i \in g^{\Phi}(\alpha)} \varphi_i \wedge \bigwedge_{-\varphi_j \in g^{\Phi}(\alpha)} \neg\varphi_j$ is consistent, as desired.

(2). Suppose that $\varphi \in |g^{\Phi}(\alpha)|$. Then, $\bigcirc\varphi \in |\alpha|$. As α is strong closed, $|\alpha|$ is closed under subformulas (recall Proposition 4.19). Thus, $\varphi \in |\alpha|$. Since α is a Φ-atom, $|\alpha| \subseteq \Phi$. Consequently, $\varphi \in \Phi$. Therefore, $|g^{\Phi}(\alpha)| \subseteq \Phi$.

(3). We now prove that $|g^{\Phi}(\alpha)|$ satisfies the closure conditions $(SC1)$-$(SC8)$. To see this, it is crucial to notice that $|\alpha|$ satisfies them as well. For details, we merely show that for $(SC7)$-$(SC8)$, and all others are routine.

$(SC7)$. Assume that $\varphi \in |g^{\Phi}(\alpha)|$ includes no operator D_X for any $X \subseteq \mathbb{V}_\Phi$. Let $Y \subseteq \mathbb{V}_\Phi$ be non-empty. We now show $\mathsf{D}_Y\varphi \in |g^{\Phi}(\alpha)|$. From $\varphi \in |g^{\Phi}(\alpha)|$, it follows that $\bigcirc\varphi \in |\alpha|$. By clause $(SC7)$, it holds that $\mathsf{D}_Y \bigcirc \varphi \in |\alpha|$. Then, as $|\alpha|$ also satisfies $(SC7)$, $\bigcirc\mathsf{D}_Y\varphi \in |\alpha|$. Immediately, $\mathsf{D}_Y\varphi \in |g^{\Phi}(\alpha)|$.

($SC8$). Suppose that $\bigcirc^m \mathsf{D}_X \bigcirc^n \varphi \in |g^{\Phi}(\alpha)|$. Let $Y \subseteq \mathbb{V}_{\Phi}$ be non-empty, and $m', n' \in \mathbb{N}$ such that $m+n = m'+n'$. We are going to prove $\bigcirc^{m'} \mathsf{D}_Y \bigcirc^{n'} \varphi \in |g^{\Phi}(\alpha)|$. It is easy to see that $\bigcirc^{m+1} \mathsf{D}_X \bigcirc^n \varphi \in |\alpha|$. Since $m+1+n = m'+1+n'$ and $|\alpha|$ satisfies ($SC8$), $\bigcirc^{m'+1} \mathsf{D}_Y \bigcirc^{n'} \varphi \in |\alpha|$. Consequently, $\bigcirc^{m'} \mathsf{D}_Y \bigcirc^{n'} \varphi \in |\alpha|$, as expected. ∎

We now prove that $\mathcal{M}^{\Phi}$ is a general relational model, i.e., its components satisfy the conditions imposed in Definition 4.10. First of all, the following holds:

Proposition 4.24: Let $\Phi \subseteq \mathcal{L}_{\mathsf{D}}$ be strong closed and $X \subseteq \mathbb{V}_{\Phi}$ be non-empty. The relation $=^{\Phi}_X$ is an equivalence relation.

Proof It is easy to see that the relation $=^{\Phi}_X$ is reflexive and symmetric. Thus, it remains to prove that it is transitive.

Let $\alpha =^{\Phi}_X \beta$ and $\beta =^{\Phi}_X \gamma$. Note that $\alpha = \varnothing$ is trivial, as it implies $\beta = \gamma = \varnothing$. We now consider the case that $\alpha \neq \varnothing$. Then, for any $D^n_X Y$, it is simple to see that:

$$+D^n_X Y \in \alpha \Leftrightarrow +D^n_X Y \in \beta \Leftrightarrow +D^n_X Y \in \gamma.$$

Also, assume that $+D^n_X Y \in \alpha$, then it is a matter of direct checking that:

$$+\bigcirc^n \mathsf{D}_Y \varphi \in \alpha \Leftrightarrow +\bigcirc^n \mathsf{D}_Y \varphi \in \beta \Leftrightarrow +\bigcirc^n \mathsf{D}_Y \varphi \in \gamma.$$

This completes the proof. ∎

Next, we have the following:

Proposition 4.25: Let $\mathcal{M}^{\Phi}$ be the Φ-model. For each Φ-atom α, $(D^n_X y)^{\alpha}$ satisfies 'Dep-Reflexivity', 'Dyn-Transitivity' and 'Determinism'.

Proof From Definition 4.22, it is easy to see that clauses M4.1 and M4.2 guarantee that $(D^n_X y)^{\alpha}$ satisfies 'Determinism' and 'Dep-Reflexivity' respectively. We still need to prove that $(D^n_X y)^{\alpha}$ has property 'Dyn-Transitivity' as well. Now assume that $\alpha \in V^{\Phi}(D^n_X Y)$ and $(g^{\Phi})^n(\alpha) \in V^{\Phi}(D^m_Y Z)$. There are different situations.

(1). Consider that $(g^{\Phi})^n(\alpha) = \varnothing$. Then, there are still two cases: $\alpha = \varnothing$ and $\alpha \neq \varnothing$.

(1.1). Consider that $\alpha = \varnothing$. When $n = 0$, it holds that $Y \subseteq X$. So, from $\varnothing \in V^{\Phi}(D^m_Y Z)$ it follows that $(g^{\Phi})^n(\alpha) \in V^{\Phi}(D^{0+m}_X Z)$, i.e., $\alpha \in V^{\Phi}(D^{n+m}_X Z)$. Also, the case is similar when $m = 0$. We now proceed to consider $n, m \neq 0$. Then, we have $X = Y = \mathbb{V}_{\Phi}$. Therefore, it holds directly that $\alpha \in V^{\Phi}(D^{n+m}_X Z)$.

(1.2). The case is that $\alpha \neq \varnothing$. As $(g^{\Phi})^n(\alpha) = \varnothing$, $D^n_X Y \notin |\alpha|$. So, from clause M4.3.2 we know that there is some $i < n$ such that $+D^i_X \mathbb{V}_{\Phi} \in \alpha$. Immediately, $\alpha \in V^{\Phi}(D^{n+m}_X Z)$.

(2). Let us move to $(g^{\Phi})^n(\alpha) \neq \varnothing$. Thus, $\alpha \neq \varnothing$. Again, there are two cases $+D^m_Y Z \in (g^{\Phi})^n(\alpha)$ and $+D^m_Y Z \notin (g^{\Phi})^n(\alpha)$.

(2.1). First, we consider $+D^m_Y Z \in (g^{\Phi})^n(\alpha)$. Then, $+\bigcirc^n D^m_Y Z \in \alpha$. So, from clause $(SC6)$, it follows $D^{m+n}_X Z \in |\alpha|$. As α is consistent, we have $+D^{m+n}_X Z \in \alpha$. By M4.3.1, $\alpha \in V^{\Phi}(D^{n+m}_X Z)$.

(2.2). Next, we proceed to show that for $+D^m_Y Z \notin (g^{\Phi})^n(\alpha)$. Then, by M4.3.2, we have $+D^i_Y \mathbb{V}_{\Phi} \in (g^{\Phi})^n(\alpha)$ for some $i < m$. By the same reasoning as in the proof of (2.1) above, but using i and $\mathbb{V}_{\Phi}$ in place of m and Z respectively, we have $+D^{i+n}_X \mathbb{V}_{\Phi} \in \alpha$. Now, with M4.3.2 we have $\alpha \in V^{\Phi}(D^{n+m}_X Z)$, as desired. ∎

Besides, we also need to prove that:

Proposition 4.26: Let $\mathcal{M}^{\Phi}$ be the Φ-model. If $\alpha =^{\Phi}_X \beta$ and $\alpha \in V^{\Phi}(D^n_X Y)$, then it holds $(g^{\Phi})^n(\alpha) =^{\Phi}_Y (g^{\Phi})^n(\beta)$ and $\beta \in V^{\Phi}(D^n_X Y)$.

Proof When $\alpha = \varnothing$, we have $\alpha = \beta = (g^{\Phi})^n(\alpha) = (g^{\Phi})^n(\beta) = \varnothing$. So, it follows that $(g^{\Phi})^n(\alpha) =^{\Phi}_Y (g^{\Phi})^n(\beta)$ and $\beta \in V^{\Phi}(D^n_X Y)$. We now proceed to show that for $\alpha \neq \varnothing$. There are also two different cases $(g^{\Phi})^n(\alpha) = \varnothing$ and $(g^{\Phi})^n(\alpha) \neq \varnothing$.

(1). $(g^{\Phi})^n(\alpha) = \varnothing$. First, we show that $(g^{\Phi})^n(\beta)$ is $\varnothing$ as well. Suppose not. Then, by Proposition 4.21 it is not hard to check that $D_{\mathbb{V}_{\Phi}} x \in |(g^{\Phi})^n(\beta)|$. So, $\bigcirc^n D_{\mathbb{V}_{\Phi}} x \in |\beta|$. Therefore, from $(SC6)$, it follows $D^n_{\mathbb{V}_{\Phi}} x \in |\beta|$. As β is consistent, $+D^n_{\mathbb{V}_{\Phi}} x \in \beta$. So, $\alpha =^{\Phi}_X \beta$ is followed by $+D^n_{\mathbb{V}_{\Phi}} x \in \alpha$. Again, using $(SC6)$, it is not hard to check that $(g^{\Phi})^n(\alpha) \neq \varnothing$, a contradiction. Immediately, $(g^{\Phi})^n(\alpha) =^{\Phi}_Y (g^{\Phi})^n(\beta)$.

Next, we show that $\beta \in V^{\Phi}(D^n_X Y)$. When $X = \mathbb{V}_{\Phi}$, it holds directly $\beta \in V^{\Phi}(D^n_X Y)$ by clause M4.1 of Definition 4.22. It remains to show that for $X \neq \mathbb{V}_{\Phi}$. From the reasoning above, $n \leqslant 1$. Also, it is easy to see that $\alpha \in V^{\Phi}(D^n_X Y)$ comes from M4.3.2, i.e., there is some $m < n$ such that $+D^m_X \mathbb{V}_{\Phi} \in \alpha$. As $\alpha =^{\Phi}_X \beta$, $+D^m_X \mathbb{V}_{\Phi} \in \beta$. So, using M4.3.2 once more, we have $\beta \in V^{\Phi}(D^n_X Y)$.

(2). Let us move to $(g^{\Phi})^n(\alpha) \neq \varnothing$.

(2.1). We first consider the case that $X = \varnothing$. Then, by clause M3 it holds trivially that $(g^{\Phi})^n(\alpha) =^{\Phi}_{\varnothing} (g^{\Phi})^n(\beta)$. Also, it is worth noting that $(g^{\Phi})^n(\alpha) \neq \varnothing$ and $\alpha \in V^{\Phi}(D^n_X Y)$ are followed by $+D^n_X Y \in \alpha$. So, from $\alpha =^{\Phi}_X \beta$ it follows that $+D^n_X Y \in \beta$. Therefore, $\beta \in V^{\Phi}(D^n_X Y)$.

(2.2). We now consider the case that $X \neq \varnothing$. By similar reasoning to the case above, we have $+D^n_X Y \in \alpha$ and $\beta \in V^{\Phi}(D^n_X Y)$. It remains to show $(g^{\Phi})^n(\alpha) =^{\Phi}_Y (g^{\Phi})^n(\beta)$, which can be achieved by two steps.

(2.2.1). Let $+D^m_Y Z \in (g^{\Phi})^n(\alpha)$. We are going to show that $+D^m_Y Z \in (g^{\Phi})^n(\beta)$. Obviously, $+ \bigcirc^n D^m_Y Z \in \alpha$. As $\bigcirc^n D^m_Y Z \in |\alpha|$, from Proposition 4.20 we have $\bigcirc^n \mathsf{D}_Y D^m_Y Z \in |\alpha|$. Now, by the consistency of α, it is easy to see $+ \bigcirc^n \mathsf{D}_Y D^m_Y Z \in \alpha$. Also, as $+D^n_X Y \in \alpha$ and $\alpha =^{\Phi}_X \beta$, it holds $+ \bigcirc^n \mathsf{D}_Y D^m_Y Z \in \beta$. Then, from clause $(SC2)$ and the consistency of β, it follows that $+ \bigcirc^n D^m_Y Z \in \beta$. By the definition of g^{Φ}, $+D^m_Y Z \in \beta$. The other direction is proved in a similar manner.

(2.2.2). Let $+D^m_Y Z, + \bigcirc^m \mathsf{D}_Z \varphi \in (g^{\Phi})^n(\alpha)$. We will prove $+ \bigcirc^m \mathsf{D}_Z \varphi \in (g^{\Phi})^n(\alpha)$. From (2.2.1) above, it holds that $+D^m_Y Z \in (g^{\Phi})^n(\beta)$. Moreover, $+ \bigcirc^{m+n} \mathsf{D}_Z \varphi \in \alpha$ and $+ \bigcirc^n D^m_Y Z \in \alpha$. Then, by the clause $(SC6)$, we have $D^{m+n}_X Z \in |\alpha|$. Recall $+D^n_X Y \in \alpha$. By the consistency of α, it follows that $+D^{m+n}_X Z \in \alpha$. Hence, from $\alpha =^{\Phi}_X \beta$ we know $+ \bigcirc^{m+n} \mathsf{D}_Z \varphi \in \beta$. Immediately, we have $+ \bigcirc^m \mathsf{D}_Z \varphi \in (g^{\Phi})^n(\beta)$. Again, the other direction is similar.

Now the proof is completed. ∎

Moreover, it holds that:

Proposition 4.27: Let $\mathcal{M}^{\Phi}$ be the Φ-model. If $\alpha =^{\Phi}_X \beta$, $(g^{\Phi})^n(\alpha) \in V^{\Phi}(P\mathbf{y})$ and $\alpha \in V^{\Phi}(D^n_X Y)$ (Y is the set of variables occurring in $\mathbf{y}$), then $(g^{\Phi})^n(\beta) \in V^{\Phi}(P\mathbf{y})$.

Proof From $(g^{\Phi})^n(\alpha) \in V^{\Phi}(P\mathbf{y})$ it follows that $+P\mathbf{y} \in (g^{\Phi})^n(\alpha)$ (recall clause M4 in Definition 4.22). Hence, $+ \bigcirc^n P\mathbf{y} \in \alpha$. Consequently, $\mathbf{td}(\alpha) \geqslant n$. By Proposition 4.21, it is simple to see that $\alpha \in V^{\Phi}(D^n_X Y)$ is followed by $+D^n_X Y \in \alpha$. Obviously, formula $\bigcirc^n P\mathbf{y}$ includes no dependence quantifiers. So, by clause $(SC7)$, it holds that $\mathsf{D}_Y \bigcirc^n P\mathbf{y} \in |\alpha|$. Then, by clause $(SC8)$, it follows that $\bigcirc^n \mathsf{D}_Y P\mathbf{y} \in |\alpha|$. Note that Y is the set of variables occurring in the tuple $\mathbf{y}$. Thus, by the consistency of α, it holds $+ \bigcirc^n \mathsf{D}_Y P\mathbf{y} \in \alpha$. Now, as $\alpha =^{\Phi}_X \beta$, from $+D^n_X Y, + \bigcirc^n \mathsf{D}_Y P\mathbf{y} \in \alpha$ we know that $+ \bigcirc^n \mathsf{D}_Y P\mathbf{y} \in \beta$. By clause $(SC2)$, it follows that $\bigcirc^n P\mathbf{y} \in |\beta|$. Since β is consistent, $+ \bigcirc^n P\mathbf{y} \in \beta$. Then, we have $+P\mathbf{y} \in (g^{\Phi})^n(\beta)$. Immediately, $g^{\Phi n}(\beta) \in V^{\Phi}(P\mathbf{y})$. ∎

Then, from Proposition 4.24-4.27, it follows immediately that:

Proposition 4.28: For any strong closed $\Phi \subseteq \mathcal{L}_{\mathsf{D}}$, the Φ-model $\mathcal{M}^{\Phi}$ is a general relational model (without the clause on universal relation $=_{\varnothing}$).

Furthermore, we have the following result:

Proposition 4.29: Given a strong closed $\Phi \subseteq \mathcal{L}_{\mathsf{D}}$, let $\mathcal{M}^{\Phi}$ be the Φ-model and $\alpha \neq \varnothing$ a Φ-atom. If $+\varphi \in \alpha$ or $-\varphi \in \alpha$, then:

$$\alpha \vDash \varphi \Leftrightarrow +\varphi \in \alpha.$$

Proof The proof goes by induction on $\varphi \in \mathcal{L}_{\mathsf{D}}$. Suppose that $+\varphi \in \alpha$ or $-\varphi \in \alpha$. The case for atoms follows from the definition of V^{Φ} directly. The cases for Boolean connectives $\neg, \wedge$ are routine. We now consider $\bigcirc\psi$ and $\mathsf{D}_X\psi$.

(1). φ is $\bigcirc\psi$. $\alpha \vDash \varphi$ iff $g^{\Phi}(\alpha) \vDash \psi$. As $+\bigcirc\psi \in \alpha$ or $-\bigcirc\psi \in \alpha$, $g^{\Phi}(\alpha) \neq \varnothing$. Then, by the inductive hypothesis, $g^{\Phi}(\alpha) \vDash \psi$ iff $+\psi \in g^{\Phi}(\alpha)$. Now, by the definition of g^{Φ}, $+\psi \in g^{\Phi}(\alpha)$ iff $+\bigcirc\psi \in \alpha$.

(2). φ is $\mathsf{D}_X\psi$. Here we consider the two directions separately. Let us begin with the easy part.

(2.1). Assume that $+\mathsf{D}_X\psi \in \alpha$. Let β be a Φ-atom such that $\alpha =_X^{\Phi} \beta$. Clearly, $\beta \neq \varnothing$. Also, $+D_X X \in \alpha$. By the definition of $=_X^{\Phi}$, it holds that $+\mathsf{D}_X\psi \in \beta$. Also, as β is consistent, we obtain $+\psi \in \beta$. By the inductive hypothesis, it follows that $\beta \vDash \psi$. Consequently, $\alpha \vDash \mathsf{D}_X\psi$.

(2.2). Suppose that $+\mathsf{D}_X\psi \notin \alpha$. Let $\Gamma = \{+D_X^n Y \mid +D_X^n Y \in \alpha\} \cup \{+\bigcirc^n \mathsf{D}_Y\varphi' \in \alpha \mid +D_X^n Y \in \alpha\}$ and $\Gamma' = \Gamma \cup \{-\psi\}$. It is easy to see that $\mathbb{V}_{\Phi} = \mathbb{V}_{|\Gamma|} = \mathbb{V}_{|\Gamma'|}$. To prove $\alpha \nvDash \mathsf{D}_X\psi$, we are going to take 2 steps:

- First, we prove that Γ' is consistent.
- Next, we construct a non-empty Φ-atom β such that $\Gamma' \subseteq \beta$ and $\alpha =_X^{\Phi} \beta$.

Given these, we have $-\psi \in \beta$. Then, by the inductive hypothesis, $\beta \nvDash \psi$. Thus, $\alpha \nvDash \mathsf{D}_X\psi$. Now let us begin to prove them.

(2.2.1). We suppose for reductio that the pseudo-atom Γ' is inconsistent. Then, as Γ is consistent, it follows that $\mathbf{I}(\Gamma) \to \psi$. Consequently, $\mathsf{D}_X\mathbf{I}(\Gamma) \to \mathsf{D}_X\psi$. Note that the identification $\mathbf{I}(\Gamma)$ is a conjunction of formulas of the forms $D_X^n Y$ and $\bigcirc^n\mathsf{D}_Y\varphi'$. Moreover, whenever $\bigcirc^n\mathsf{D}_Y\varphi'$ is a conjunct of $\mathbf{I}(\Gamma)$, $D_X^n Y$ is also a conjunct of $\mathbf{I}(\Gamma)$. Thus, we have $\mathbf{I}(\Gamma) \leftrightarrow \mathsf{D}_X\mathbf{I}(\Gamma)$. So, $\mathbf{I}(\Gamma) \to \mathsf{D}_X\psi$. However, $+\mathsf{D}_X\psi \notin \alpha$ is followed by $-\mathsf{D}_X\psi \in \alpha$. Immediately, α is inconsistent, a contradiction. The first step is completed.

(2.2.2). Next, we construct a Φ-atom β such that $\Gamma' \subseteq \beta$ and $\alpha =_X^{\Phi} \beta$. Note that it is possible that $\Gamma' \subseteq \alpha$ (i.e., $-\psi \in \alpha$), as Γ' is consistent. If so, α itself is exactly what we need: by the inductive hypothesis, we have $\alpha \nvDash \psi$. Consequently, we obtain $\alpha \nvDash \mathsf{D}_X\psi$,

as $\alpha =^{\Phi}_{X} \alpha$. In what follows, we consider the other case that $-\psi \notin \alpha$. Since $-\mathsf{D}_X\psi \in \alpha$, we have $\psi \in |\alpha|$. Therefore, from $-\psi \notin \alpha$ it follows that $+\psi \in \alpha$.

Consider the following enumeration of formulas $D^m_X Y$ for some $m \in \mathbb{N}$ and $Y \subseteq \mathbb{V}_{\Phi}$ such that $D^m_X Y \in Cl(\mathbf{I}(\Gamma'))$ and $-D^m_X Y \in \alpha$:

Enumeration 1: $\quad D^{n_1}_X Y_1,\ D^{n_2}_X Y_2,\ \cdots,\ D^{n_j}_X Y_j$

Note that this enumeration is finite, as $Cl(\mathbf{I}(\Gamma'))$ is finite. Also, formulas in the enumeration cannot be of the form ψ, since $+\psi \in \alpha$.

We now construct j pseudo-atoms as follows:

$$\Gamma'_1 = \begin{cases} \Gamma' \cup \{+D^{n_1}_X Y_1\} & \textit{if } \mathbf{I}(\Gamma') \to D^{n_1}_X Y_1 \\ \Gamma' \cup \{-D^{n_1}_X Y_1\} & \textit{otherwise} \end{cases}$$

$$\vdots$$

$$\Gamma'_j = \begin{cases} \Gamma'_{j-1} \cup \{+D^{n_j}_X Y_j\} & \textit{if } \mathbf{I}(\Gamma'_{j-1}) \to D^{n_j}_X Y_j \\ \Gamma'_{j-1} \cup \{-D^{n_j}_X Y_j\} & \textit{otherwise} \end{cases}$$

As Γ' is consistent, from the construction it is easy to know that all these j pseudo-atoms are also consistent. Except the consistency of these pseudo-atoms, another important reason that we construct them in such a way is: we aim to make the final Γ'_j to contain as few formulas $+D^m_X Y$ with $-D^m_X Y \in \alpha$ as possible. Essentially, we have the following:

Claim 1. For all formulas $D^{n_{j'}}_X Y_{j'}$ in Enumeration 1, we have $-D^{n_{j'}}_X Y_{j'} \in \Gamma'_j$.

Proof Let us now prove the claim. We first consider $D^{n_1}_X Y_1$. Suppose that $+D^{n_1}_X Y_1 \in \Gamma'_1$. Then, by the construction, it holds that $\mathbf{I}(\Gamma) \wedge \neg\psi \to D^{n_1}_X Y_1$. So, $\mathbf{I}(\Gamma) \wedge \neg D^{n_1}_X Y_1 \to \psi$. Consequently, $\mathsf{D}_X\mathbf{I}(\Gamma) \wedge \mathsf{D}_X \neg D^{n_1}_X Y_1 \to \mathsf{D}_X\psi$. Similar to that of (2.2.1), $\mathbf{I}(\Gamma) \leftrightarrow \mathsf{D}_X\mathbf{I}(\Gamma)$. Also, as $\mathsf{D}_X \neg D^{n_1}_X Y_1 \leftrightarrow \neg D^{n_1}_X Y_1$, it holds $\mathbf{I}(\Gamma) \wedge \neg D^{n_1}_X Y_1 \to \mathsf{D}_X\psi$. However, $\mathbf{I}(\Gamma) \wedge \neg D^{n_1}_X Y_1$ is a conjunct of $\mathbf{I}(\alpha)$, which implies $+\mathsf{D}_X\psi \in \alpha$, a contradiction.

Afterwards, with the help of $-D^{n_1}_X Y_1 \in \Gamma'_1$, we can show $-D^{n_2}_X Y_2 \in \Gamma'_2$. Repeating the reasoning, we can finally prove that $-D^{n_j}_X Y_j \in \Gamma'_j$. Therefore, for all formulas $D^{n_{j'}}_X Y_{j'}$ in Enumeration 1, it holds that $-D^{n_{j'}}_X Y_{j'} \in \Gamma'_j$. ∎

Now, it follows that for all $D^n_X Y$,

$$+D^n_X Y \in \Gamma'_j \text{ iff } +D^n_X y \in \alpha.$$

Next, let the following be an enumeration of formulas $\bigcirc^m \mathsf{D}_Z\psi'$ for some $m \in \mathbb{N}$ and $Z \subseteq \mathbb{V}_{\Phi}$ such that $\bigcirc^m \mathsf{D}_Z\psi' \in Cl(\mathbf{I}(\Gamma'))$, $+D_X^m Z \in \Gamma'_j$ and $-\bigcirc^m \mathsf{D}_Z\psi' \in \alpha$:

Enumeration 2: $\bigcirc^{m_1}\mathsf{D}_{Z_1}\psi'_1,\ \bigcirc^{m_2}\mathsf{D}_{Z_2}\psi'_2,\ \cdots,\ \bigcirc^{m_i}\mathsf{D}_{Z_i}\psi'_i$

Again, Enumeration 2 is finite, and all formulas occurring in it cannot be ψ. Based on the pseudo-atom Γ'_j, we now construct i pseudo-atoms:

$$\Gamma'_{j(1)} = \begin{cases} \Gamma'_j \cup \{+\bigcirc^{m_1} \mathsf{D}_{Z_1}\psi'_1\} & \textit{if } \mathbf{I}(\Gamma'_j) \to \bigcirc^{m_1}\mathsf{D}_{Z_1}\psi'_1 \\ \Gamma'_j \cup \{-\bigcirc^{m_1} \mathsf{D}_{Z_1}\psi'_1\} & \textit{otherwise} \end{cases}$$

$$\vdots$$

$$\Gamma'_{j(i)} = \begin{cases} \Gamma'_{j(i-1)} \cup \{+\bigcirc^{m_i} \mathsf{D}_{Z_i}\psi'_i\} & \textit{if } \mathbf{I}(\Gamma'_{j(i-1)}) \to \bigcirc^{m_i}\mathsf{D}_{Z_i}\psi'_i \\ \Gamma'_{j(i-1)} \cup \{-\bigcirc^{m_i} \mathsf{D}_{Z_i}\psi'_i\} & \textit{otherwise} \end{cases}$$

By the construction, it is easy to see that $\Gamma'_{j(i)}$ is consistent. Also, $Cl(\mathbf{I}(\Gamma'))$ may contain other formulas not occurring in Enumeration 1 or Enumeration 2. Anyway, we can always obtain a Φ-atom $\beta \supseteq \Gamma'_{j(i)}$ by adding to $\Gamma'_{j(i)}$ suitable forms $+\psi'$ or $-\psi'$ of all other formulas $\psi' \in Cl(\mathbf{I}(\Gamma'))$. For pseudo-atom $\Gamma'_{j(i)}$, we claim the following:

Claim 2. For all $\bigcirc^{m_{i'}}\mathsf{D}_{Z_{i'}}\psi'_{i'}$ in Enumeration 2, we have $-\bigcirc^{m_{i'}} \mathsf{D}_{Z_{i'}}\psi'_{i'} \in \Gamma'_{j(i)}$.

Proof Let us first consider formula $\bigcirc^{m_1}\mathsf{D}_{Z_1}\psi'_1$. If $+\bigcirc^{m_1} \mathsf{D}_{Z_1}\psi'_1 \in \Gamma'_{j(1)}$, then it holds that $\mathbf{I}(\Gamma'_j \setminus \{-\psi\}) \wedge \neg\psi \to \bigcirc^{m_1}\mathsf{D}_{Z_1}\psi'_1$. Equivalently, $\mathbf{I}(\Gamma'_j \setminus \{-\psi\}) \wedge \neg \bigcirc^{m_1} \mathsf{D}_{Z_1}\psi'_1 \to \psi$. Consequently, $\mathsf{D}_X\mathbf{I}(\Gamma'_j \setminus \{-\psi\}) \wedge \mathsf{D}_X\neg \bigcirc^{m_1} \mathsf{D}_{Z_1}\psi'_1 \to \mathsf{D}_X\psi$. Similar to that of (2.2.1), it holds that $\mathsf{D}_X\mathbf{I}(\Gamma'_j \setminus \{-\psi\}) \leftrightarrow \mathbf{I}(\Gamma'_j \setminus \{-\psi\})$. Also, from $+D_X^{m_1} Z_1 \in \Gamma'_j$ (recall the definition of Enumeration 2), it follows that $\neg \bigcirc^{m_1} \mathsf{D}_{Z_1}\psi'_1 \leftrightarrow \mathsf{D}_X\neg \bigcirc^{m_1} \mathsf{D}_{Z_1}\psi'_1$. Thus, we obtain $\mathbf{I}(\Gamma'_j \setminus \{-\psi\}) \wedge \neg \bigcirc^{m_1} \mathsf{D}_{Z_1}\psi'_1 \to \mathsf{D}_X\psi$. However, $\mathbf{I}(\Gamma'_j \setminus \{-\psi\}) \wedge \neg \bigcirc^{m_1} \mathsf{D}_{Z_1}\psi'_1$ is a conjunct of $\mathbf{I}(\alpha)$, and so $+\mathsf{D}_X\psi \in \alpha$, a contradiction.

Next, suppose that for all $i' \leqslant i-1$, $-\bigcirc^{m_{i'}} \mathsf{D}_{Z_{i'}}\psi'_{i'} \in \Gamma'_{j(i)}$. We now prove that $-\bigcirc^{m_i} \mathsf{D}_{Z_i}\psi'_i \in \Gamma'_i$. If not, then $\mathbf{I}(\Gamma'_j \setminus \{-\psi\}) \wedge \neg\psi \wedge \neg \bigcirc^{m_1} \mathsf{D}_{Z_1}\psi'_1 \wedge \cdots \wedge \neg \bigcirc^{m_{i-1}} \mathsf{D}_{Z_{i-1}}\psi'_{i-1} \to \bigcirc^{m_i}\mathsf{D}_{Z_i}\psi'_i$. Note that $+D_X^{m_{j'}} Z_{j'} \in \Gamma'_j$ for all $1 \leqslant j' \leqslant i$. Similar to the basic case above, we can also obtain $+\mathsf{D}_X\psi \in \alpha$, which contradicts to our assumption. ∎

With Claim 1 and Claim 2, it is easy to check that $\alpha =_X^{\Phi} \beta$. Since $-\psi \in \beta$, $+\psi \notin \beta$. By the inductive hypothesis, $\beta \not\models \psi$. Therefore, $\alpha \not\models \mathsf{D}_X\psi$. ∎

Now we proceed to show that the fragment of logic DFD without $D^n_\varnothing y$ and $D_\varnothing \varphi$ enjoys the finite model property w.r.t. general relational models:

Theorem 4.8: The fragment of logic DFD without $D^n_\varnothing y$ and $D_\varnothing \varphi$ has the finite model property w.r.t. general relational models, i.e., if a formula $\varphi \in \mathcal{L}_{\mathsf{D}}$ is satisfiable, then it is satisfied in a finite general relational model.

Proof Let $\varphi \in \mathcal{L}_{\mathsf{D}}$ be satisfiable. Consider the $Cl(\varphi)$-model $\mathcal{M}^{Cl(\varphi)}$, which is a finite general relational model. Since φ is satisfiable, there exists some $Cl(\varphi)$-atom α such that $+\varphi \in \alpha$. By Proposition 4.29, it follows that $\mathcal{M}^{Cl(\varphi)}, \alpha \vDash \varphi$, as desired. ∎

As a consequence, it holds that:

Theorem 4.9: The fragment of the logic DFD without $D^n_\varnothing y$ and $D_\varnothing \varphi$ is decidable.

Chapter 5 Information-sensitive diffusion in social networks

5.1 Introduction

So far, we have explored dynamic dependence in a very general sense. As stated in Chapter 4, although our original motivation originated from social interactions in game scenarios, dynamic dependence with a time delay is a basic notion that need not be restricted to games. In this chapter, we consider one important concrete non-game like scenario: diffusion of opinions (or behaviors, or products) in a community. For simplicity, we drop the explicit temporal dimension, but we will add stepwise update modalities instead.

Dynamic dependence in diffusion arises since behaviours or opinions are often influenced by others in social life. When considering to buy a Huawei phone, my decision may depend on whether there is a large population around me using it. In such scenarios, how a trend spreads through a population depends on two factors: (a) the structure of the population, and (b) how easy it is for agents to get influenced by others.

From a logical perspective, almost all mentioned proposals restrict themselves to social networks with only a singular kind of social relation (representing, e.g., *following* or *friendships*) and agents in these networks are influenced just by their direct neighbors.[1] This makes sense, as direct neighbors intuitively stand for agents who are around us. However, there are also many exceptions. For instance, as suggested by Baltag et al. (2019b), when deciding whether or not to support a revolution, a crucial factor is if a big enough part of the total population, not only our neighbors, is supporting the revolution. Inspired by these phenomena, in this chapter we develop a logical framework highlighting the difference between direct neighbors and sources of influence.

Moreover, in doing so, we are led to take on board another crucial aspect of realistic social scenarios: namely, the epistemic fact that agents can only make use of the information that is available to them. Taking this further, the agents studied in this chapter can communicate with each other, and as we shall see, the neighborhood relation itself is then

1 Baltag et al. (2019b) discuss *prediction updates* which allow an agent to adopt the opinion or behavior in question if she knows that a large enough proportion of her direct neighbors will adopt it (even if those neighbors have not adopted it now). This enables agents to make the best use of their information, but the underlying assumption is still that agents are influenced by their *direct* neighbors.

exactly the communication channel.

Outline of the chapter. In Section 5.2, we introduce some preliminary notions, including social networks, threshold models and updates of behaviors. Next, a basic logical proposal is explored in Section 5.3 to reason about the diffusion of behaviors in networks. Afterwards, we enrich the basic notion of threshold models with an epistemic dimension in Section 5.4, which also discusses their updates induced by different types of operations and provides us with a new logic, including its language and semantics. Then, Section 5.5 develops a complete Hilbert-style calculus for the dynamic-epistemic logic. Finally, we end this chapter by Section 5.6 on conclusion and further directions.

5.2 Preliminary notions

In this section, let us introduce some preliminaries, mainly including the notions of social networks, threshold models and the policy on the update of behaviors.

Roughly, a social network can be represented as a graph with many binary relations, where nodes can be treated as agents and relations intuitively stand for different kinds of social relationships among them. Depending on the features of relations to be modeled, many further restrictions might be imposed. Say, 'friendships' are often assumed to be *symmetric*. Thus, when an agent a is a friend of another agent b, then b is a friend of a as well. Moreover, in our setting, the relation is also assumed to be *reflexive*. In contrast, we do not have the same assumptions on the relation of 'following': for instance, when buying medicines, we follow the suggestions of doctors, but not the other way around. In this chapter, we consider the networks including these two kinds of relations, and restrict ourselves to finite graphs. Sometimes we also employ neighbors and influence relation respectively for friends and following. Formally, the definition of social networks is as follows:

Definition 5.1: A *social network* is a tuple $\langle \mathcal{A}, \mathcal{N}, \mathcal{I} \rangle$ where $\mathcal{A} \neq \varnothing$ is a finite set of agents, the function $\mathcal{N} : \mathcal{A} \to \mathcal{P}(\mathcal{A})$ assigns a set $\mathcal{N}(a) \subseteq \mathcal{A}$ to each $a \in \mathcal{A}$ such that

- $a \in \mathcal{N}(a)$ (Reflexivity)
- $a \in \mathcal{N}(b)$ if and only if $b \in \mathcal{N}(a)$ (Symmetry)

and the function $\mathcal{I}\colon \mathcal{A} \to \mathcal{P}(\mathcal{A})$ assigns a set $\mathcal{I}(a) \subseteq \mathcal{A}$ to each $a \in \mathcal{A}$ such that

- $\mathcal{I}(a) \neq \varnothing$ (Seriality).

Intuitively, for an agent $a \in \mathcal{A}$, $\mathcal{N}(a)$ is the set of her direct neighbors or friends,

and $\mathcal{I}(a)$ denotes those she follows. Generally, $\mathcal{I}(a)$ is different from $\mathcal{N}(a)$. Also, as indicated by the clause $\mathcal{I}(a) \neq \varnothing$, we are only interested in the situations where every agent does follow some agents, since otherwise agents never change their behaviors, which is irrelevant to our current discussion on diffusion in chapter. Based on the notion of social networks, we are able to introduce the following:

Definition 5.2: A *threshold model* is a tuple $\mathcal{M} = \langle \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta \rangle$, where $\langle \mathcal{A}, \mathcal{N}, \mathcal{I} \rangle$ is a social network, $B \subseteq \mathcal{A}$ is a behavior and $\theta \in [0, 1]$ is a uniform adoption threshold.

In the definition, we identify a behavior with its extension, i.e., those agents who have already adopted it. Besides, the threshold θ intuitively represents how easy it is for the agents in $\mathcal{A}$ to be affected by others. For simplicity, in the chapter we just consider the case that the threshold is uniform, i.e., all agents have the same threshold. However, it is instructive to notice that this can be relaxed very easily. Given a threshold model, we can also calculate the spread of the behavior in question as follows:

Definition 5.3: Let $\mathcal{M} = \langle \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta \rangle$ be a threshold model. Its *update* is $\mathcal{M}' = \langle \mathcal{A}, \mathcal{N}, \mathcal{I}, B', \theta \rangle$, where $B' := \{a \in \mathcal{A} \mid \frac{|\mathcal{I}(a) \cap B|}{|\mathcal{I}(a)|} \geqslant \theta\}$ and for all sets A, $|A|$ refers to its cardinality.

Therefore, the resulting model is the same as the original one, except that the behavior now is B' other than B, consisting of those agents whose influence sets includes a large enough proportion of agents having adopted the behavior before the update. It is worth noting that the way of updating has several distinguishing features. Some of them are as follows:

- First, it is determined whether an agent should adopt or unadopt the behavior after a round of updating.
- Next, the policy indicates that all agents are forced to adopt or unadopt the behavior by the *fact* of others' behavior. Thus, the underlying assumption is that the information of agents is always available to each other.
- Finally, the neighbor relation $\mathcal{N}$ among agents does *not* play any role in the process of diffusion.

The first one does reflect the essence of the notion of threshold models. But the second one looks too strong to be realistic, as it leaves no room for uncertainty. However, in many real-life situations, we may only be able to act in accordance with the information

available to us. Therefore, in the reminder of the chapter, one of our main goals is to consider the process of diffusion in an epistemic setting, in which, as we will see, the neighbor relation is not superfluous, but an important ingredient that has an explicit influence on diffusion processes.

Remark 5.1: Our policy on update is in line with the spirit of the so-called *Susceptible-Infected-Susceptible models*, in which agents are not only permitted to adopt the behavior in question, but also to *unadopt* it. Moreover, there are also many other manners to give the update. Say, a more restrictive way is to replace B with $\{a \in \mathcal{A} \mid \frac{|\mathcal{I}(a)\cap B|}{|\mathcal{I}(a)|} > \theta\}$. Compared with the one given in Definition 5.3, it requires that an agent should adopt the behavior only when the proportion of those having adopted it in her influence set is strictly *larger* than the threshold θ. Also, one can consider it as $\{a \in \mathcal{A} \mid \frac{|\mathcal{I}(a)\cap B|}{|\mathcal{I}(a)|} > \theta\} \cup \{a \in \mathcal{A} \mid \frac{|\mathcal{I}(a)\cap B|}{|\mathcal{I}(a)|} = \theta, a \in B\}$, as suggested by Baltag et al. (2019b). Then, in the new setting, when the proportion of those having adopted the behavior in the influence set of an agent is exactly the threshold, the agent need not change her current stance. More generally, it is also interesting to explore policies fitting with the *Susceptible-Infected models* that only allow agents to adopt the behavior in question, but not to unadopt it. In such a setting, the new set of behavior can be defined as, e.g., $\{a \in \mathcal{A} \mid \frac{|\mathcal{I}(a)\cap B|}{|\mathcal{I}(a)|} \geqslant \theta\}$ or $\{a \in \mathcal{A} \mid \frac{|\mathcal{I}(a)\cap B|}{|\mathcal{I}(a)|} > \theta\}$. The former one was studied by Baltag et al. (2019b), and all others also deserve to be studied in future.

5.3 A dynamic logic for updates of threshold models

Before moving to the more complicated setting, let us first introduce a dynamic logic for modelling the notion of threshold models and their dynamics. Essentially, the logic is just a simple adaptation of that of (Baltag et al., 2019b), but it still deserves to be introduced a bit detailed: except giving us a formal tool to reason about diffusion processes, the logic itself is also a foundation of our epistemic framework.

Definition 5.4: Let $\mathcal{A}$ be a nonempty, finite set. Atomic propositions are given by $\{I_{ab} \mid a, b \in \mathcal{A}\} \cup \{N_{ab} \mid a, b \in \mathcal{A}\} \cup \{\beta_a \mid a \in \mathcal{A}\}$. *Language* $\mathcal{L}_b$ is given by the following grammar:

$$\varphi ::= I_{ab} \mid N_{ab} \mid \beta_a \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid [\mathsf{A}]\varphi$$

The abbreviations $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined as usual. For each natural number $n \in \mathbb{N}$, we denote by $[\mathsf{A}]^n\varphi$ the n-th iteration of $[\mathsf{A}]$. When $n = 0$, $[\mathsf{A}]^n\varphi$ is just φ.

Intuitively, formula N_{ab} states that *agents a and b are neighbors/friends*, I_{ab} reads *a is influenced by b*, β_a means *agent a has already adopted the behavior*, and finally, formula $[\mathsf{A}]\varphi$ expresses that φ *is the case after a round of adoption-update* (or more precisely, φ *holds after all agents update their behavior with the rule given in Definition 5.3 at the same time*). Their truth conditions are as follows:

Definition 5.5: Let $\mathcal{M} = \langle \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta \rangle$ be a threshold model and $\varphi \in \mathcal{L}_b$. The *semantics for* $\mathcal{L}_d$ is inductively defined in the following:

$$\begin{aligned}
\mathcal{M} \vDash I_{ab} &\quad \Leftrightarrow \quad b \in \mathcal{I}(a) \\
\mathcal{M} \vDash N_{ab} &\quad \Leftrightarrow \quad b \in \mathcal{N}(a) \\
\mathcal{M} \vDash \beta_a &\quad \Leftrightarrow \quad a \in B \\
\mathcal{M} \vDash \neg\varphi &\quad \Leftrightarrow \quad \mathcal{M} \nvDash \varphi \\
\mathcal{M} \vDash \varphi \wedge \psi &\quad \Leftrightarrow \quad \mathcal{M} \vDash \varphi \text{ and } \mathcal{M} \vDash \psi \\
\mathcal{M} \vDash [\mathsf{A}]\varphi &\quad \Leftrightarrow \quad \mathcal{M}' \vDash \varphi
\end{aligned}$$

where $\mathcal{M}'$ is the update of $\mathcal{M}$ produced in the way given by Definition 5.3.

Thus, the resulting logic is essentially just a propositional logic. Although it is simple, perhaps surprisingly it is powerful enough to give us many useful characterizations that looks much strong. One example is as follows:

$$\beta_{I(a)\geqslant\theta} := \bigvee_{\{\mathcal{G}\subseteq\mathfrak{I}\subseteq\mathcal{A} \mid \frac{|\mathcal{G}|}{|\mathfrak{I}|}\geqslant\theta\}} (\bigwedge_{b\in\mathfrak{I}} I_{ab} \wedge \bigwedge_{b\notin\mathfrak{I}} \neg I_{ab} \wedge \bigwedge_{b\in\mathcal{G}} \beta_b)$$

Intuitively, the set $\mathfrak{I}$ used in the formula aims to capture the influence set of agent a, and $\mathcal{G}$ standards for the set of β-agents in $\mathcal{I}(a)$. Hence, $\beta_{I(a)\geqslant\theta}$ is true in $\mathcal{M}$ if, and only if, the proportion of agents who have adopted the behavior currently in a's influence set is equal or larger than the threshold θ. Now, with the abbreviation, Table 5.1 presents a proof system, written as $\mathbf{LTM}_\theta$, for logic of the threshold models with threshold θ.

The network axioms indicates the features of social networks, i.e., the influence relation is serial and friendships are both reflexive and symmetric. Also, the second part is the *recursion axioms* for the adoption operator $[\mathsf{A}]$, reducing dynamic formulas into static ones. Maybe the most interesting principle is $[\mathsf{A}]$-β, which suggests that an agent a would adopt the behavior after a round of adoption-update if, and only if, before the update there is already a subset $\mathcal{G}$ of $\mathcal{I}(a)$ such that $\mathcal{G} \subseteq B$ and $\frac{|\mathcal{G}|}{|\mathcal{I}(a)|} \geqslant \theta$.

Let $\theta \in [0, 1]$ and $\mathfrak{M}_\theta$ the class of threshold models with the threshold θ. Now, let us proceed to show the following:

Table 5.1 Proof system $\mathbf{LTM}_\theta$. Subscripts a, b are arbitrary over $\mathcal{A}$.

I	The classical propositional logic
II	Network axioms:
N-Reflexivity	N_{aa}
N-Symmetry	$N_{ab} \leftrightarrow N_{ba}$
I-Seriality	$\bigvee_{b\in\mathcal{A}} I_{ab}$
III	Recursion axioms for $[\mathsf{A}]$:
$[\mathsf{A}]$-N	$[\mathsf{A}]N_{ab} \leftrightarrow N_{ab}$
$[\mathsf{A}]$-I	$[\mathsf{A}]I_{ab} \leftrightarrow I_{ab}$
$[\mathsf{A}]$-β	$[\mathsf{A}]\beta_a \leftrightarrow \beta_{I(a)\geqslant\theta}$
$[\mathsf{A}]$-$\neg$	$[\mathsf{A}]\neg\varphi \leftrightarrow \neg[\mathsf{A}]\varphi$
$[\mathsf{A}]$-$\wedge$	$[\mathsf{A}](\varphi \wedge \psi) \leftrightarrow [\mathsf{A}]\varphi \wedge [\mathsf{A}]\psi$
IV	Inference rule:
Nec. $[\mathsf{A}]$	From φ, infer $[\mathsf{A}]\varphi$

Theorem 5.1: The calculus $\mathbf{LTM}_\theta$ is sound and complete w.r.t. $\mathfrak{M}_\theta$.

Proof *Soundness:* We merely show that the axiom $[\mathsf{A}]$-β is valid, and all others are routine. Let $\mathcal{M} \in \mathfrak{M}_\theta$. Then, we have the following sequence of equivalences: $\mathcal{M} \vDash [\mathsf{A}]\beta_a$ iff $\mathcal{M}' \vDash \beta_a$ iff $a \in B' = \{a \in \mathcal{A} \mid \frac{|\mathcal{I}(a)\cap B|}{|\mathcal{I}(a)|} \geqslant \theta\}$. It is worth noting that $a \in B'$ iff $\beta_{I(a)\geqslant\theta}$ is true in $\mathcal{M}$. So, $\mathcal{M} \vDash [\mathsf{A}]\beta \leftrightarrow \beta_{I(a)\geqslant\theta}$.

Completeness: The static part of the logic is just a propositional logic, whose completeness w.r.t. social networks is easy to prove. Furthermore, the completeness of the whole logic can be established by making use of the recursion axioms. ∎

5.4 A dynamic-epistemic logic for diffusion

As suggested by Definition 5.3, whether an agent should adopt the behavior in question depends on the fact that if her influence set contains a large enough population with the behavior. However, in our real-life situations, if or not an agent *knows* the fact matters. So, in this part we will augment our existing proposal with an epistemic dimension, to express the uncertainty about the facts. Moreover, based on the knowledge of agents, the new setting enables them to communicate with their friends and update their relationships with others and their behaviors. Before moving to the semantic part, let us first fix our language, which is a straightforward extension of language $\mathcal{L}_b$:

Definition 5.6: Let $\mathcal{A}$ be a nonempty, finite set. Again, atoms are given by $\{I_{ab} \mid a, b \in \mathcal{A}\} \cup \{N_{ab} \mid a, b \in \mathcal{A}\} \cup \{\beta_a \mid a \in \mathcal{A}\}$. The *language* $\mathcal{L}_e$ is defined as follows:

$$\varphi ::= I_{ab} \mid N_{ab} \mid \beta_a \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid P_a^n\varphi \mid [\mathsf{A}]\varphi \mid [C]\varphi$$

where $n \in \mathbb{N}$ is a natural number. For brevity, by $K_a\varphi$ we denote $P_a^0\varphi$. Also,we employ $\widehat{P}_a^n\varphi$ for $\neg P_a^n\neg\varphi$, and $\widehat{K}_a\varphi$ for $\neg K_a\neg\varphi$.

So, compared with language $\mathcal{L}_b$, now we have two additional components: $P_a^n\varphi$ and $[C]\varphi$. For any $n \in \mathbb{N}$, P_a^n is called *potential knowledge operator*, and $P_a^n\varphi$ reads *agent a potentially knows φ (from those agents that are n-reachable*[1] *from a)*. Also, as we will see, K is exactly the usual knowledge operator: $K_a\varphi$ states *a knows φ*. Moreover, $[\mathsf{A}]\varphi$ means that *after a round of adoption update, φ is the case*. Formula $[C]\varphi$ shows that φ *holds after communication*. Moreover, for each $n \in \mathbb{N}$, we define N^n such that:

- $N_{ab}^0 := \top$ if $a = b$, and $N_{ab}^0 := \bot$ if $a \neq b$.
- $N_{ab}^{n+1} := \bigvee_{c \in \mathcal{A}} (N_{ac}^1 \wedge N_{cb}^n)$.

5.4.1 Epistemic threshold models and their updates

First of all, let us introduce the notion of 'epistemic threshold models':

Definition 5.7: An *epistemic threshold model with threshold θ* (ETM_θ) $\mathcal{M}$ is a tuple $\langle \mathcal{W}, \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta, \{\sim_a\}_{a \in \mathcal{A}} \rangle$ such that:

- $\mathcal{W}$ is a finite, non-empty set of possible worlds or states.
- $\mathcal{A}$ is a finite, non-empty set of agents.
- $\mathcal{N} : \mathcal{W} \to (\mathcal{A} \to \mathcal{P}(\mathcal{A}))$ assigns a neighborhood $\mathcal{N}(w)(a)$ to each $a \in \mathcal{A}$ in each $w \in \mathcal{W}$ s.t. $a \in \mathcal{N}(w)(a)$ and $a \in \mathcal{N}(w)(b)$ iff $b \in \mathcal{N}(w)(a)$.
- $\mathcal{I} : \mathcal{W} \to (\mathcal{A} \to \mathcal{P}(\mathcal{A}))$ assigns a nonempty influence set $\mathcal{I}(w)(a)$ to each $a \in \mathcal{A}$ in each $w \in \mathcal{W}$ s.t. $\mathcal{I}(w)(a) \neq \emptyset$.
- $B : \mathcal{W} \to \mathcal{P}(\mathcal{A})$ assigns to each $w \in \mathcal{W}$ a set $B(w)$ of agents who have adopted the behavior.
- $\theta \in [0, 1]$ is a uniform adoption threshold.
- $\sim_a \subseteq \mathcal{W} \times \mathcal{W}$ is an equivalence relation for each agent $a \in \mathcal{A}$.

The definition may look complicated, but the underlying intuition is rather simple: an ETM is just some threshold models (in the sense of Definition 5.2) with a fixed set $\mathcal{A}$ of

1 For the meaning of 'n-reachable', see Definition 5.8.

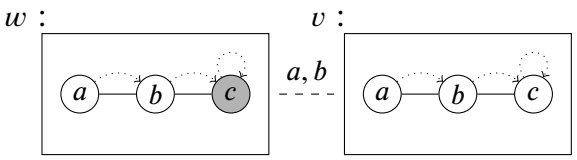


Figure 5.1 An ETM $\mathcal{M}$. The model consists of two possible worlds $\mathcal{W} = \{w, v\}$. The dashed line labelled with agents between possible worlds represents that the agents cannot distinguish the situations (we omit reflexive and transitive links when representing indistinguishability relations, and reflexive links when representing friendships). In each possible world, the friendships are represented by the undirected links and a dotted link from an agent to another indicates that the latter influences the former. Also, $B(w) = \{c\}$, i.e., only agent c adopted the behavior in world w, and $B(v) = \emptyset$. Now, let us assume that $\theta = 1$ and the actual case is w. Then, with the policy given by Definition 5.3, agent b should adopt the behavior in one round of adoption-update, and then a will also adopt it finally. However, neither of them know that.

agents, which are connected by the equivalence relations $\sim_{a \in \mathcal{A}}$ expressing the uncertainty of agents. For an example, see Figure 5.1.

Furthermore, we can also impose some natural restrictions on ETM_θ. For instance, in the chapter we always assume that:

Agents always know their own behaviors, and their friends and influence sets.

Precisely, the assumption is characterized by the clause that for all worlds $w, v \in \mathcal{W}$ and agents $a, b \in \mathcal{A}$, if $w \sim_a v$, then $a \in B(w)$ iff $a \in B(v)$, $b \in \mathcal{N}(w)(a)$ iff $b \in \mathcal{N}(v)(a)$, and $b \in \mathcal{I}(w)(a)$ iff $b \in \mathcal{I}(v)(a)$.

Essentially, the restriction endows agents with the 'ability' to eliminate some uncertainty about different situations. However, this by no means says that agents *only* know these facts: some of them may know more by accident. For any $\theta \in [0, 1]$, we denote by $\mathcal{E}_\theta$ *the class of epistemic threshold models with threshold* θ *and satisfying the above restriction*.

Also, for any ETM, we introduce the following notion of '$\mathcal{N}$*-distance*' (for short, *distance*) between agents:

Definition 5.8: Let $\mathcal{M} = \langle \mathcal{W}, \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta, \{\sim_a\}_{a \in \mathcal{A}} \rangle$ be an ETM, $w \in \mathcal{W}$ and $a \in \mathcal{A}$. For every natural number $1 \leqslant n \in \mathbb{N}$, define $\mathcal{N}^n : \mathcal{W} \to \mathcal{A} \to \mathcal{P}(\mathcal{A})$ in the following:

- $\mathcal{N}^1(w)(a) = \mathcal{N}(w)(a)$
- $\mathcal{N}^{n+1}(w)(a) = \mathcal{N}^n(w)(a) \cup \{b \in \mathcal{A} \mid \exists c \in \mathcal{N}^n(w)(a) \text{ and } b \in \mathcal{N}(w)(c)\}$

So, $b \in \mathcal{N}^n(w)(a)$ indicates that in possible world w, agent b can be reached from agent a at most in n-steps via relation $\mathcal{N}$. In this case, we say that agent b is *n-reachable* from agent a in w.

Let $\mathcal{M} = \langle \mathcal{W}, \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta, \{\sim_a\}_{a\in\mathcal{A}}\rangle$ be an ETM, $a \in \mathcal{A}$ and $w \in \mathcal{W}$. For any $v \in \mathcal{W}$ and $n \in \mathbb{N}$, we define that $w \sim^n_{\langle a\rangle} v$ iff $w \sim_a v$ and $w \sim_b v$ for all $b \in \mathcal{N}^n(w)(a)$. Here it is worth noting that each $\sim^n_{\langle a\rangle}$ is still an equivalence relation. In particular, when $n = 0$, $\sim^n_{\langle a\rangle}$ is identical to $\sim_a$.

We now have enough background to introduce the truth conditions for the static part of our language $\mathcal{L}_e$, and details are as follows:

$$\begin{array}{rcl}
\mathcal{M}, w \vDash \beta_a & \Leftrightarrow & a \in B(w) \\
\mathcal{M}, w \vDash N_{ab} & \Leftrightarrow & b \in \mathcal{N}(w)(a) \\
\mathcal{M}, w \vDash I_{ab} & \Leftrightarrow & b \in \mathcal{I}(w)(a) \\
\mathcal{M}, w \vDash \neg\varphi & \Leftrightarrow & \mathcal{M}, w \nvDash \varphi \\
\mathcal{M}, w \vDash \varphi \wedge \psi & \Leftrightarrow & \mathcal{M}, w \vDash \varphi \text{ and } \mathcal{M}, w \vDash \psi \\
\mathcal{M}, w \vDash P^n_a\varphi & \Leftrightarrow & \text{for all } v \in \mathcal{W}, \text{ if } w \sim^n_{\langle a\rangle} v \text{ then } \mathcal{M}, v \vDash \varphi
\end{array}$$

By the truth condition for P^n_a, it is easy to see:

$$\mathcal{M}, w \vDash K_a\varphi \quad \Leftrightarrow \quad \text{for all } v \in \mathcal{W}, \text{ if } w \sim_a v, \text{ then } \mathcal{M}, v \vDash \varphi$$

which illustrates that K_a is an **S5**-operator characterized by the equivalence relation $\sim_a$ directly. Moreover, essentially all P^n_a are **S5**. But compared with the notion of knowledge, potential knowledge plays a more general role in our setting: knowledge is always a kind of potential knowledge, but potential knowledge is not necessarily knowledge. More generally, for any $\varphi \in \mathcal{L}_e$, $a \in \mathcal{A}$ and $m, n \in \mathbb{N}$, formula $P^n_a\varphi \to P^{n+m}_a\varphi$ is valid, but $P^{n+m}_a\varphi \to P^n_a\varphi$ may fail. To see the latter, consider the following:

Example 5.1: Let us consider again the ETM depicted in Figure 5.1. We have $\mathcal{N}^1(w)(a) = \{a, b\}$ and $\mathcal{N}^2(w)(a) = \{a, b, c\}$. Now, it is not hard to see that at state w formula $P^1_a\beta_c$ is false while $P^2_a\beta_c$ is true. So, it holds immediately that $\mathcal{M}, w \nvDash P^2_a\beta_c \to P^1_a\beta_c$. Therefore, the schema $P^{n+m}_a\varphi \to P^n_a\varphi$ is not valid.

Also, in terms of potential knowledge, the structure $\mathcal{N}$ of friendships in a network is crucial, and the knowledge of a's friends definitely contributes to her potential knowledge. For instance, for any $1 \leqslant m, n \in \mathbb{N}$, it holds that

$$N_{ab} \wedge K_a\varphi \wedge K_b\psi \to P^m_a(\varphi \wedge \psi) \wedge P^n_b(\varphi \wedge \psi)$$

Now, we proceed to discuss the dynamic part of our framework. First of all, with the richer devices ETM, an important question is: how do agents update their behavior now?

As suggested by Baltag et al. (2019b), there are at least two natural ways to do so. One of them is that:

Definition 5.9: Let $\mathcal{M} = \langle \mathcal{W}, \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta, \{\sim_a\}_{a\in\mathcal{A}}\rangle$ be an ETM. Its *adoption update* is $\mathcal{M}' = \langle \mathcal{W}, \mathcal{A}, \mathcal{N}, \mathcal{I}, B', \theta, \{\sim_a\}_{a\in\mathcal{A}}\rangle$, where

$$B'(w) := \{a \in \mathcal{A} \mid \forall v \sim_a w : \tfrac{|\mathcal{I}(v)(a)\cap B(v)|}{|\mathcal{I}(v)(a)|} \geqslant \theta\}.$$

Therefore, whether an agent a adopts or unadopts the behavior is based on her *de dicto* knowledge on the behavior of agents in her influence set: she does not need to know exactly who have already adopted the behavior, and what matters is that if or not in all possible situations her influence set contains a large enough fraction of the agents with the behavior. In the reminder of the chapter, we will focus on this clause of update. Let us now introduce the interpretation of operator $[\mathsf{A}]$:

$$\mathcal{M}, w \vDash [\mathsf{A}]\varphi \quad \Leftrightarrow \quad \mathcal{M}', w \vDash \varphi \text{ where } \mathcal{M}' \text{ is given by Definition 5.9.}$$

Other kinds of update. As stated already, with Definition 5.9, if or not an agent in the network adopts the behavior depends on her *de dicto* knowledge on the actions of the agents in her influence set. Different from this, we can also introduce other kinds of update policies. For instance, Baltag et al. (2019b) also suggested the the policy involving agents' *de re* knowledge: $B'(w) = \{a \in \mathcal{A} \mid \frac{|\{b\in\mathcal{A}|\forall v\sim_a w: b\in\mathcal{I}(v)(a)\cap B(v)\}|}{|\{b\in\mathcal{A}|\forall v\sim_a w: b\in\mathcal{I}(v)(a)\}|} \geqslant \theta\}$, which requires proportion of agents in $\mathcal{I}(w)(a)$ known by a to be β is large enough. This policy is definitely stronger than that of Definition 5.9.

Finally, to show the semantics completely, it remains to present the truth condition for the communication operator $[C]$. For this, we need to specify how to update models with the operator. To make the spread of the behavior in question as efficient as possible, a desirable way is to share all their information with all friends. Usually, dynamic epistemic logics make use of explicit formulas to specify the contents being communicated. However, this does not always work. As noted by Baltag and Smets (2020),

- Depending on the expressivity of the language, their might be no formula in the language that can capture all knowledge of an agent.
- Even when there exists such a formula, the total sum of an agent's knowledge can only be expressed by a huge formula. Also, in a purely syntactic approach, the order of announcements matter: previously expressible information may become inexpressible after another announcement, which may prevent the full resolution of distributed knowledge (cf. van Benthem, 2006).

Therefore, we employ such a communication operator not specifying explicit contents communicated, which aims to capture that all agents share all they know with their friends simultaneously. Precisely, the communication update can be formally defined as follows:

Definition 5.10: Let $\mathcal{M} = \langle \mathcal{W}, \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta, \{\sim_a\}_{a\in\mathcal{A}}\rangle$ be an ETM. The resulting model from the *communication update* is $\mathcal{M}^C = \langle \mathcal{W}, \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta, \{\sim'_a\}_{a\in\mathcal{A}}\rangle$, where $\sim'_a := \sim^1_{\langle a\rangle}$ for any agent $a \in \mathcal{A}$.

Thus, the relation $\mathcal{N}$ among agents now is essentially the channel of their communication. For brevity, given an ETM $\mathcal{M}$, we write $\mathcal{M}^{nC}$ for the resulting model after $1 \leqslant n \in \mathbb{N}$ rounds of communication. Here it is worth spending a few words on the peculiar features of the notion of communication given above.

- After a round of communication, what an agent knows essentially consists of the distributed knowledge of herself and her friends before the communication.
- In terms of the information flow, agents' 'locations' w.r.t. relation $\mathcal{N}$ in a network play a significant role (Carrington, 2013). For instance, given two agents a, b, if all a's neighbors are also neighbors of b, i.e., $\bigwedge_{c\in\mathcal{A}} (N_{ac} \to N_{bc})$, then what a potentially knows from one-reachable friends is also potentially known by b from one-reachable friends, i.e.,

$$\bigwedge_{c\in\mathcal{A}} (N_{ac} \to N_{bc}) \to (P_a\varphi \to P_b\varphi).$$

- Possible worlds and agents are finite in an ETM, so there exists a class of the longest $\mathcal{N}$-sequences. Let **SEQ** be such a sequence. In terms of knowledge flow, we can reach a fixed point at most $|\mathbf{SEQ}|$ rounds of communication, i.e., $\mathcal{M}^{|\mathbf{SEQ}|C} = \mathcal{M}^{(|\mathbf{SEQ}|+1)C}$.

With this definition, we now can define the truth condition for $[C]$:

$$\boxed{\mathcal{M}, w \vDash [C]\varphi \quad \Leftrightarrow \quad \mathcal{M}^C, w \vDash \varphi}$$

Immediately, it is not hard to check that the following formula concerning the relation of K_a, P_a and $[C]$ is valid:

$$[C]K_a\varphi \leftrightarrow P_a[C]\varphi$$

which states that after communication agent a knows φ if, and only if, agent a potentially knows φ after communication.

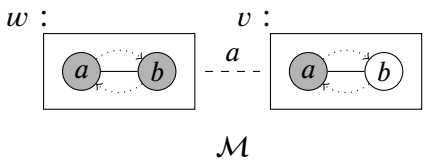


Figure 5.2 A case showing that communication influences the result of diffusion. Let $\mathcal{M}$ be an ETM such that the threshold $\theta = 1$, $\mathcal{I}(w)(a) = \mathcal{I}(v)(a) = \{b\}$, $\mathcal{I}(w)(b) = \mathcal{I}(v)(b) = \{a\}$. Agent b knows the actual world is w, where both a and b have adopted the behavior. However, agent a cannot distinguish w from v. In world v, a has adopted the behavior, but b is not. In such a setting, a would unadopt the behavior in the next round of adoption update (and both a and b will unadopt the behavior in next two rounds of update). But, if we let they communicate first, then a will know the world w is the actual case, and agents a, b will keep the property forever.

It might be worth noting that, the communication between friends may change the result of adoption update, which is determined by the behavior of those in agents' influence sets, especially when agents in the influence set of an agent can also be reached from the agent through the her friends. That is, it may happen that some friends tell us useful information on experts. For an example for this, see Figure 5.2.

Digression. Let $\mathcal{M} = \langle \mathcal{W}, \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta, \{\sim_a\}_{a\in\mathcal{A}} \rangle$ be an ETM and $a \in \mathcal{A}$. As stated, the communication update introduced in Definition 5.10 intuitively expresses that all agents tell all they know to their neighbors at the same time. Different from this, we can also define a version of stepwise communication $[!a]\varphi$ expressing that φ is the case after a tell all she knows to her neighbors. The update induced by $[!a]$ is obtained by replacing $\{\sim_a\}_{a\in\mathcal{A}}$ with $\{\sim'_a\}_{a\in\mathcal{A}}$ defined as follows:

- For a: $w \sim'_a v$ iff $w \sim_a v$
- For $b \in \mathcal{N}(w)(a)$: $w \sim'_b v$ iff $w \sim_a v$ and $w \sim_b v$
- For any $b \notin \mathcal{N}(w)(a) \cup \{a\}$: $w \sim'_b v$ iff $w \sim_b v$.[1]

It is worth noting that $[C]$ is not the same as the case that every agents tell all they know to their neighbors one by one: the structure of $\mathcal{N}$ in our setting matters.

5.5 Axiomatization

At the final of this chapter, we show a complete Hilbert-style calculus $\mathbf{LET}_\theta$ for the logic. See Table 5.2 for its details.

1 The preliminary version of this operator, operating on the same models as those of the public announcement logic PAL, is introduced in Dr. Alexandru Baltag's course 'Dynamic Epistemic Logics' at the ILLC, who called it 'Tell All You Know' modality.

Table 5.2 Proof system $\mathbf{LET}_\theta$, where $n, m \in \mathbb{N}$ and $a, b \in \mathcal{A}$.

I	The classical propositional logic
II	Network axioms: the part **II** of Table 5.1
III	Knowledge-network axioms:
Known neighbors	$N_{ab} \to K_a N_{ab}$
Known influence	$I_{ab} \to K_a I_{ab}$
Known behavior	$\beta_a \to K_a \beta_a$
IV	Potential knowledge axioms:
K-P_a^n	$P_a^n(\varphi \to \psi) \to (P_a^n \varphi \to P_a^n \psi)$
T-P_a^n	$P_a^n \varphi \to \varphi$
4-P_a^n	$P_a^n \varphi \to P_a^n P_a^n \varphi$
5-P_a^n	$\widehat{P}_a^n \varphi \to P_a^n \widehat{P}_a^n \varphi$
V	Interaction axioms:
N-P^n	$N_{ab} \wedge P_a^n \varphi \to P_b^{n+1} \varphi$
N^m-P^n	$\bigwedge_{c \in \mathcal{A}} (N_{ac}^n \to N_{bc}^m) \to (P_a^n \varphi \to P_b^m \varphi)$
VI	Recursion axioms for $[\mathsf{A}]$:
$[\mathsf{A}]$-N	$[\mathsf{A}] N_{ab} \leftrightarrow N_{ab}$
$[\mathsf{A}]$-I	$[\mathsf{A}] I_{ab} \leftrightarrow I_{ab}$
$[\mathsf{A}]$-β	$[\mathsf{A}] \beta_a \leftrightarrow K_a \beta_{I(a) \geqslant \theta}$
$[\mathsf{A}]$-$\neg$	$[\mathsf{A}] \neg \varphi \leftrightarrow \neg [\mathsf{A}] \varphi$
$[\mathsf{A}]$-$\wedge$	$[\mathsf{A}](\varphi \wedge \psi) \leftrightarrow [\mathsf{A}] \varphi \wedge [\mathsf{A}] \psi$
$[\mathsf{A}]$-P^n	$[\mathsf{A}] P_a^n \varphi \leftrightarrow P_a^n [\mathsf{A}] \varphi$
VII	Recursion axioms for $[C]$:
$[C]$-N	$[C] N_{ab} \leftrightarrow N_{ab}$
$[C]$-I	$[C] I_{ab} \leftrightarrow I_{ab}$
$[C]$-β	$[C] \beta_a \leftrightarrow \beta_a$
$[C]$-$\neg$	$[C] \neg \varphi \leftrightarrow \neg [C] \varphi$
$[C]$-$\wedge$	$[C](\varphi \wedge \psi) \leftrightarrow [C] \varphi \wedge [C] \psi$
$[C]$-P^n	$[C] P_a^n \varphi \leftrightarrow P_a^{n+1} [C] \varphi$
VIII	Inference rules:
Nec.P^n	From φ, infer $P_a^n \varphi$, for any $a \in \mathcal{A}$
Nec.$[\mathsf{A}]$	From φ, infer $[\mathsf{A}] \varphi$
Nec.$[C]$	From φ, infer $[C] \varphi$

The static part of $\mathbf{LET}_\theta$ consists of the classical propositional logic, network axioms, knowledge-network axioms, potential knowledge axioms, interaction axioms and inference rule Nec.P^n. The network axioms are the same as those in Table 5.1. The knowledge-network axioms show agents know their own behaviors, their neighbors and those agents who influence them. Moreover, potential knowledge axioms characterize the fact that each $\sim^n_{\langle a\rangle}$ is an equivalence relation. Also, both axioms N-P^n and N^m-P^n are principles illustrating the interactions between the relation $\mathcal{N}$ and potential knowledge. Intuitively, the former states that if a and b are friends and a potentially knows φ from those n-reachable from herself, then b potentially knows φ from those who are $(n+1)$-reachable. Additionally, principles N^m-P^n expresses that if all agents that are n-reachable from a are m-reachable from b, then all a potentially knows from those that are n-reachable is also potentially known by b from those are m-reachable.

Let us now briefly comment on the principles of the dynamic part. Formulas [A]-N and [A]-I show that the adoption update does not affect the structure of the network captured by $\mathcal{N}$ or $\mathcal{I}$. Principle [A]-β illustrates that one agent a becomes β after a round of adoption update if, and only if, before the update, in each possible situation considered by a, the proportion of agents having adopted in her influence set is always above or equal to the threshold. The principle [A]-P^n shows that for any $n \in \mathbb{N}$, [A] and P^n_a are commutative. In terms of operator $[C]$, its recursion axioms involving Boolean cases are similar to those of [A]. The principle $[C]$-P^n states that after a round of communication, agent a potentially knows φ from those n-reachable from her if, and only if, agent a potentially knows from those $(n+1)$-reachable from her that after a round of communication φ is the case.

Theorem 5.2: Let $\theta \in [0,1]$ and $n \in \mathbb{N}$. For any $\varphi \in \mathcal{L}_e$, it holds that:

$$\mathcal{C}_\theta \vDash \varphi \Leftrightarrow \mathbf{LET}_\theta \vdash \varphi.$$

Proof *Soundness.* Let $\mathcal{M} = \langle \mathcal{W}, \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta, \{\sim_a\}_{a\in\mathcal{A}}\rangle$ be an ETM of $\mathcal{C}_\theta$, $w \in W$, $a, b \in \mathcal{A}$ and $n \in \mathbb{N}$. It is simple to see that all network axioms are valid. Also, the knowledge-network axioms hold directly by the semantics and the assumption that all agents know their neighbors and the agents who influence them. Besides, all $\sim^n_{\langle a\rangle}$ are still equivalence relations, so it is not hard to see all the potential knowledge axioms are valid. Furthermore, it can also be checked straightforward that the interaction axioms are valid. Now we consider the validity of recursion axioms. It is not to see that all recursion axioms involving $\neg$ and $\wedge$ are obvious. Let us begin with those for [A].

As illustrated by Definition 5.9, the operator [A] only affects the extension of B, hence validity of formulas [A]-N and [A]-I are trivial. Let $\mathcal{M}'$ be the adoption update of $\mathcal{M}$. We now prove [A]-β is valid:

$$\begin{aligned}
\mathcal{M}, w \vDash [\mathsf{A}]\beta_a &\Leftrightarrow \mathcal{M}', w \vDash \beta_a \\
&\Leftrightarrow a \in \{a \in \mathcal{A} \mid \forall v \sim_a w : \tfrac{|\mathcal{I}(v)(a) \cap B(v)|}{|\mathcal{I}(v)(a)|} \geqslant \theta\} \\
&\Leftrightarrow \text{for each } v \in \mathcal{W},\ w \sim_a v \text{ entails } \tfrac{|\mathcal{I}(v)(a) \cap B(v)|}{|\mathcal{I}(v)(a)|} \geqslant \theta \\
&\Leftrightarrow \text{for each } v \in \mathcal{W}, \text{ if } w \sim_a v, \text{ then there exist two sets} \\
&\quad \mathfrak{I}(= \mathcal{I}(v)(a)) \text{ and } \mathcal{G}(= B(v) \cap \mathcal{I}(v)(a)) \text{ s.t. } \tfrac{|\mathcal{G}|}{|\mathfrak{I}|} \geqslant \theta \\
&\Leftrightarrow \text{for each } v \in \mathcal{W}, \text{ if } w \sim_a v, \text{ then } \mathcal{M}, v \vDash \beta_{I(a) \geqslant \theta} \\
&\Leftrightarrow \mathcal{M}, w \vDash K_a \beta_{I(a) \geqslant \theta}
\end{aligned}$$

Next, it is easy to see that [A]-P^n is also valid:

$$\begin{aligned}
\mathcal{M}, w \vDash [\mathsf{A}]P_a^n \varphi &\Leftrightarrow \mathcal{M}', w \vDash P_a^n \varphi \\
&\Leftrightarrow \text{for all } v \in \mathcal{W} \text{ s.t. } w \sim^n_{\langle a \rangle} v,\ \mathcal{M}', v \vDash \varphi \\
&\Leftrightarrow \text{for all } v \in \mathcal{W} \text{ s.t. } w \sim^n_{\langle a \rangle} v,\ \mathcal{M}, v \vDash [\mathsf{A}]\varphi \\
&\Leftrightarrow \mathcal{M}, v \vDash P_a^n [\mathsf{A}]\varphi
\end{aligned}$$

Now we move to considering those for $[C]$. Since the communication update does not affect the structures $\mathcal{N}, \mathcal{I}$ of networks or behaviors of agents, the validity of $[C]$-N, $[C]$-I and $[C]$-β holds immediately by the semantics. We prove the case for $[C]$-P^n. Let $\mathcal{M}^C = \langle \mathcal{W}, \mathcal{A}, \mathcal{N}, \mathcal{I}, B, \theta, \{\sim'_a\}_{a \in \mathcal{A}} \rangle$ be the communication update of $\mathcal{M}$. Then, we have the following sequence of equivalences:

$$\begin{aligned}
\mathcal{M}, w \vDash [C]P_a^n \varphi &\Leftrightarrow \mathcal{M}^C, w \vDash P_a^n \varphi \\
&\Leftrightarrow \text{for all } v \in \mathcal{W} \text{ s.t. } w \sim'^n_{\langle a \rangle} v,\ \mathcal{M}^C, v \vDash \varphi \\
&\Leftrightarrow \text{for all } v \in \mathcal{W} \text{ s.t. } w \sim^{n+1}_{\langle a \rangle} v,\ \mathcal{M}^C, v \vDash \varphi \\
&\Leftrightarrow \text{for all } v \in \mathcal{W} \text{ s.t. } w \sim^{n+1}_{\langle a \rangle} v,\ \mathcal{M}, v \vDash [C]\varphi \\
&\Leftrightarrow \mathcal{M}, w \vDash P_a^{n+1} [C]\varphi
\end{aligned}$$

Moreover, it is not hard to see that the validity of $\mathcal{L}_e$-formulas is invariant under the inference rules. So, we conclude that $\mathbf{LET}_\theta$ is sound.

Completeness. By the recursion axioms, it can be shown that for any $\varphi \in \mathcal{L}_e$, there exists a static formula of $\psi \in \mathcal{L}_e$ such that $\mathbf{LET}_\theta \vdash \varphi \leftrightarrow \psi$. Therefore, the completeness of the logic follows if we can show the completeness of the static part. Essentially, the proof is not trivial, but it can be shown by a simple adaptation of the techniques developed by *iterated access logic* (Carrington, 2013). Therefore, we omit the details here. ∎

5.6 Summary and future work

Summary. In the chapter, based on threshold models, we studied a specific form of dynamic dependence in social interactions, i.e., the diffusion of behaviors or opinions among agents. Two formal frameworks were proposed to analyze the phenomena.

Our first proposal was built directly on the notions of threshold models and their updates, and we provided a complete proof system and discussed interesting variants. The approach was essentially motivated by Baltag et al. (2019b), though with some noteworthy differences. For instance, the social networks we were interested in contained different sorts of social relations. However, as indicated, not all relations played significant roles in processes of diffusion, but this did not mean that distinguishing them made no sense: in fact, it was the scenarios investigated that were still too simple.

Next, after we moved to realistic scenarios in which agents' epistemic states mattered, things became much more interesting: all social relations now made their own contribution to the evolution of behaviors or opinions. To handle these scenarios precisely, we enriched our models with a notion of potential knowledge and ways of communication, and analyzed their influence on the diffusion of behaviors. Technically, a complete Hilbert-style calculus was presented for the resulting richer logic.

Further directions. The notions and results presented in this chapter suggest many open problems. A number of these resemble the challenges identified in Baltag et al. (2019b). For instance, it is meaningful to consider other kinds of policies to update behaviors, and more general dynamic logics should accommodate these. Besides, there are other directions to be explored. For instance, the social relationships we considered in this chapter are static, but it is natural to also consider their dynamics. Say, agents might make new friends (cf. e.g., Smets and Velázquez-Quesada, 2017) or remove agents from their influence set, as suggested by our work in Chapter 2. Additionally, as agents in our setting are allowed to unadopt behaviors, one natural empirical phenomenon to be explored is *oscillations* of behaviors in dynamical systems (van Benthem, 2015). Finally, one could also introduce agents' beliefs instead of just their knowledge, and also, allow a wider range of communicative acts affecting knowledge beliefs.

Yet further relevant research questions of a broader nature will be found in the concluding chapter of this dissertation.

Chapter 6 Conclusions and further directions

6.1 Conclusions

In this dissertation, we have explored social interactions between agents from a logical point of view.

The first part, consisting of chapters 2 and 3, was concerned with multi-agent interactions where agents are drastically at odds, to the extend that they change the environment in which their interaction takes place. Our tool for modeling such settings were graph games, which can be modified to fit various concrete scenarios. For describing winning strategies and other notions relevant to these games, we introduced a natural logical formalism which connects to standard modal and dynamic logics, now with additional modalities describing the effects of changing the models on which semantic interpretation takes place. This logic can be seen as the core calculus of valid reasoning about social interactions with environmental change. In chapter 2, we showed in particular that the logic of localized agents shares many features and properties with standard modal logics, but that the dynamics of environment change has a complexity which is reflected in the undecidability of the logic. The system also generated further questions of independent interest beyond the particular case of sabotage graph games.

Next, in Chapter 3, we studied what needs to happen when an abstract logic of graph change is taken to the concrete setting of learning/teaching scenarios. Even the simple example of guiding and correcting a student seeking to establish a mathematical proof involved many additional features, such as the importance of a history and the existence of different kinds of graph change for 'correction' verses 'warning'. We gave a concrete formal model for modeling such aspects, and determined the resulting logic, which can be viewed as an extension of the system of reasoning in Chapter 2, allowing us to see what additional reasoning principles govern such concrete scenarios. We believe that the resulting framework would be a natural addition to existing models in formal learning theory (Gierasimczuk, 2010; Kelly, 1996).

In a second part of the dissertation, we moved away from details of graph games to just focus on the phenomenon of dependence between actions. This may be seen as another face of logical analysis: identifying broad features of a phenomenon (in this case,

the strategic bonds that arise in groups engaging in shared action), and determining their core principles at a very abstract level.

In Chapter 4, we presented a logical analysis of dependence over time in dynamical system, an abstraction out of specific games where we just study transitions in joint states for all agents. Extending a recent decidable modal logic of abstract static (instantaneous) dependence presented in (Baltag and van Benthem, 2021b), we added flow of time, and showed that the resulting core logic of dynamic dependencies playing over time is still axiomatizable, and indeed, a sizeable part of it was shown to be decidable. This system may be considered as a sort of basic logic for temporal dependencies in social systems.

Going further, we also took into account the fact that dynamical systems, a widely used tool in many settings, from physics to computer science and formal philosophy, usually come with a topology on their state space, allowing us to bring mathematical notions like approximation and continuity to bear on epistemic and action structure. More generally, we also believe that this framework might be a natural extension of an existing area of 'dynamic topological logic' (Kremer and Mints, 2007), developed originally for analyzing the foundations of dynamical systems theory.[1]

The abstraction move toward dynamical systems also means that we now have a mathematical framework that is neutral toward the intuitive distinction between agents with 'high rationality' (Skyrms, 1990; van Benthem et al., 2021b), who deliberate, seek information and decide consciously, and agents with 'low rationality' who merely follow some hard-wired rule, perhaps biologically encoded, or as part of their software design. In our final Chapter 5, we considered one instance of these, namely threshold models for opinion formation, originally coming from sociology, but now also increasingly used in logical studies of opinion formation in large groups. This move is natural because the methodology of modal and dynamic logics still applies here (Baltag et al., 2019b; Christoff and Hansen, 2013, 2015; Liu et al., 2014; Seligman et al., 2011; Shi, 2021). We made a concrete case study of adding one crucial extra dimension in realistic social scenarios, that manifests itself as knowledge in the high-rationality realm, and as information in the low-rationality realm. We provided richer models capturing this, and found a logic showing the resulting surplus in basic reasoning over and above the bare systems studied in Chapter 4.

1 Finding a base logic of dynamical systems need not have one unique answer. For a solution that is prima face different from ours, using notions from Domain Theory, cf. the recent ILLC dissertation (Hornischer, 2021) which can handle neural networks and similar structures.

6.2 Further directions

This dissertation is an exploration of some new perspectives on interacting agents. It provides some answers to the questions posed in the Introduction, but as always, more questions came to light in our investigation. These come in different kinds. Some concern obvious continuations of our results, and we start by summarizing these chapter by chapter.

Technical continuation and outreach opportunities by chapter.

In chapters 2 and 3 forming Part I, we introduced two kinds of graph games $\mathsf{S_dG}$ and CLG as well as their matching logics $\mathsf{S_dML}$ and CLL. These logics can describe the relevant winning positions for both players in given finite graphs, and we determined many of their semantic and computational meta-properties. However, equally natural questions remained open. We did not provide *sound and complete Hilbert-style calculi*, though we believe this may be done using the techniques recently developed in (van Benthem et al., 2021a). Also, since both logics are not closed under substitution, a feature found with many dynamic-epistemic logics, one would also want to axiomatize the *schematic validities*, and determine their complexity, cf. Holliday et al. (2013) on solving this for public announcement logic. Next, our languages suggests a lot of syntactic fine-structure, starting with the limited set of formulas needed for representing the basic properties of our graph games: such *fragments* of our logics remain to be studied. On the other hand, there is also an issue of extension. Our are not expressive enough to determine *generic winning conditions* across models: for this, we need to extend our languages with fixpoint operators (cf. e.g., Kozen, 1983). Finally, taking the broader perspective on social interaction in (van Benthem and Liu, 2020), there is also the issue of how our logics interface with more general *logics of games* that tend to have much more structure, down to defining preferences, goals, equilibria and other game-theoretic features that we have left aside (van Benthem, 2014).

Chapters 4 and 5 in Part II provided logics for dynamic dependence of actions in social settings. The former chapter studied the notion at a very abstract level, in terms of behavior of variables over time in dynamical systems, while the latter explored an embodiment in social reality, viz. diffusion of behaviors or opinions among agents. The results in Chapter 4 included completeness and the decidability of a fragment of the logic, but one major problem we left open was the decidability of the whole logical system. We believe that the answer is positive, but had only an approach to offer that still needs to be validated

in detail. Another obvious desideratum would be the addition of a true future operator allowing us to reason about eventual behavior of a dynamical system, instead of just the local step-by-step dynamics. A further set of issues has to do with the topological dimension of our results. Our logics can be seen as an extension of standard modal logic with a topological semantics to include explicit reasoning about continuous functions (Baltag and van Benthem, 2021a), but can this analogy be made precise and fruitful? Finally, a junction between our approach and results with the existing body of work on dynamic topological logic seems a natural step to take, but it remains to be made. And another natural linkage would be with logical studies of causal reasoning (Halpern, 2016; Ibeling and Icard, 2020; Xie, 2020).

As for the concrete case study in Chapter 4, an obvious task remaining is a systematic comparison with other proposed modal logics for threshold models, (Baltag et al., 2019b; Christoff, 2016; Christoff and Grossi, 2017; Christoff and Hansen, 2015; Christoff et al., 2016; Rendsvig, 2014). Another natural issue is whether we should not add a component of *influence*, and let behaviors have effects on changes in influence relations (Shi, 2021)? But there are also interesting issues of comparison with the preceding chapter. One would be to add an explicit temporal logic component to our analysis of opinion formation. But on the other hand, threshold rules can also be seen as generalizing the very notion of dependence in the earlier chapter. The behavior of a variable is no longer fixed by that of a fixed set of other variables (its governing 'authority'), but by a majority vote among the other variables. What are the basic properties of this alternative notion, which seems to considerably generalize standard functional dependence?

Filling the gaps between our chapters and general logical methodology.

Our chapters represent different case studies, published in different venues for different purposes. But with them in place, a number of connection issues emerge. For instance, how does the emphasis on extensive *sequential* games in Part I fit in more detail with the emphasis on dependence, dynamical systems and *simultaneous* action in Part II? We have made some observations here, but this would merit much more attention. For instance, dynamical systems unfold one particular strategy profile for players, but the essence of the strategic situation in graph games and their corresponding modal logics might be seen as one of free choice. Also, while graph games naturally fit with *classical game theory* (Osborne and Rubinstein, 1994), (where they are a very special case), dynamical systems

fit best with *evolutionary game theory* (Hofbauer and Sigmund, 1998), and the relation between these two perspectives merits more discussion.

Another significant issue is that dynamical systems do not attach special significance to the two-agent case, arbitrary finite *groups* of agents are taken on board from the start. But this raises issues of playing graph games with larger groups of agents, and the resulting coalitional structure that have not been investigated at all.

Next, while we presented the step from games to dynamical systems as perhaps an evident and preferred abstraction, this ignores existing attempts in this direction with a very different slant. In particular, the logics of powers of players in games found in (van Benthem, 2014; van Benthem et al., 2019; van Benthem and Klein, 2020) focus on abstract *powers* of players for influencing results of the game, in the psirit of modal neighborhood semantics (Pacuit, 2017). How do these relate to the approach taken here?

One connection issue that we have already noted is the pervasive role of *knowledge* in realistic social scenarios. But we have only explored this in one case study, of dynamical threshold systems for opinion formation. Adding epistemic aspects explicitly to all issues studied in this thesis is something that remains to be done.

Connected to this, there is the role of *rationality* as a hidden assumption behind much of our analysis. To be sure, our different topics suggested a contrast, or a meeting ground, between high rationality and low rationality. But this philosophical dimension has not been taken further in this dissertation, even though it is a natural issue to raise. For instance, graph games might just as well be used, precisely because they are so simple and basic, to study the behavior of agents that do not display classical rationality, or perhaps more realistically, of studying the interplay of agents endowed with less or more features of classical rationality.

Other, still broader, questions concern our logical methodology. On the whole, this dissertation is semantic in nature, and the backdrop to our results is logical model theory. It would be of interest to see whether there is a more purely proof-theoretic approach to the core reasoning about the aspects of interactive agency studied here: environment change and dependence, perhaps extending the dialogical approach (Keiff, 2011) or the category-theoretic one of Abramsky (1995). Another desideratum would be a closer linkage between the logical style of analyzing social scenarios, and those offered by game theory: classical for high-rationality agents and evolutionary game theory for low-rationality agents.

Extending empirical coverage and questioning conceptual choices.

Finally, there is always the issue of taking on board more aspects of reality in our models. We have mentioned several instances already in our concrete case studies. In fact, about every concrete case study one can think of will bring in more notions. There are also some general desiderata coming out of this, as we already noted. For instance, in social settings, surely, in addition to looking at how individuals interact, there is the issue of how to model the behavior of *groups*. Also, social life may be said to essentially involve a mixture of dependence and *independence*, a notion we have ignored, even though the modal framework that we have adopted in Part II has also been used for giving definitions of independence and reasoning about it. We are not going to give more examples here, since finding new aspects not covered by proposed logical models is an easy game to play for which the reader does not need our guidance.

Our next observation is one of the potential impact of ideas from philosophy on studies like those in this dissertation. Here is one such issue, which might also lead us to question some choices we have made. We believe that the (sometimes benign) tension between *individuals* and *groups* is essential to understanding social life. But the way this would show in our logics is rather crude. Groups are usually just treated as sets of individuals in standard epistemic and dynamic logics. But clearly, a group is much more than a set. Now there have been many interesting attempts at formulating this surplus, making groups 'sets plus ...": cf. Paterson (2018); Shi and Wang (2021). But one direction that we see is more radical, namely dropping the use of sets altogether, and metaphysically reconceptualizing the notion of a group. What we have in mind here are *mereological theories* of extended entities with parthood relations, cf. (Varzi, 2019) for a brief introduction. As it happens, in a line of research separate from this dissertation, we have recently proposed new modal logics of mereological structure (Li and Wang, 2021), and we believe that taking this mereological perspective to the topics of this dissertation might be highly worthwhile follow-up project.

Finally, there is a general sort of tension in logical studies of real phenomena, which is also quite visible in this dissertation. On the one hand, logical studies of social reality tend to take on board more and more details, thereby creating two sorts of problems. One is that these details have often been studied already in the separate behavioral sciences, so a legitimate question arises of what does a logical analysis really adds. The art is to

know where to stop. The other problem is the challenge, or even paradox, of complexity. The more expressive and detailed we make our logics, the more complex, in general, their systems of validities, so in order to get closer to reality, we make the logical tools more and more complicated, endangering the motivation for using logic in the first place. Given this, we have also emphasized the other virtue of logical analysis, namely going firmly in the opposite direction, and abstracting from lost of realistic details to get at some high-level essence that supports a perspicuous, and hopefully not too complex, base theory of social features such as dependence of behavior.

We do not offer a fail-safe solution to the dilemma of the two directions of abstraction and concretization here: what to do and where to go may remain essentially a matter of good taste. Whether this dissertation has succeeded in striking the right balance is a question we must leave to our readers.

Bibliography

Abramsky, S. (1995). Semantics of interaction. Lecture notes, Department of Computer Science, University of Edinburgh, UK.

Andréka, H., Németi, I., and van Benthem, J. (1998). Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic*, 27:217–274.

Areces, C. (2007). Hybrid logics: The old and the new. In Arrazola, X. and Larrazabal, J., editors, *Proceedings of LogKCA '07*, page 15–29.

Areces, C., Fervari, R., and Hoffmann, G. (2012). Moving arrows and four model checking results. In Ong, L. and de Queiroz, R., editors, *Logic, Language, Information and Computation (WoLLIC '2012)*, volume 7456 of *Lecture Notes in Computer Science*, pages 142–153.

Areces, C., Fervari, R., and Hoffmann, G. (2015). Relation-changing modal operators. *Logic Journal of the IGPL*, 23:601–627.

Areces, C., Fervari, R., Hoffmann, G., and Martel, M. (2016). Relation-changing logics as fragments of hybrid logics. In Cantone, D. and Delzanno, G., editors, *Proceedings of the Seventh International Symposium on Games, Automata, Logics and Formal Verification*, volume 226 of *Electronic Proceedings in Theoretical Computer Science*, pages 16–29. Open Publishing Association.

Areces, C., Fervari, R., Hoffmann, G., and Martel, M. (2018). Satisfiability for relation-changing logics. *Journal of Logic and Computation*, 28:1143–1470.

Areces, C., Figueira, D., Figueira, S., and Mera, S. (2011). The expressive power of memory logics. *The Review of Symbolic Logic*, 4:290–318.

Areces, C. and ten Cate, B. (2007). Hybrid logics. In Blackburn, P., van Benthem, J., and Wolter, F., editors, *Handbook of Modal Logic*, pages 821–868. Elsevier.

Arenas, F. G. (1999). Alexandroff spaces. *Acta Mathematica Universitatis Comenianae*, 68:17–25.

Armstrong, M. A. (1983). *Basic Topology*. Springer-Verlag.

Artemov, S., Davoren, J., and Nerode, A. (1997). Modal logics and topological semantics for hybrid systems. Technical report, Mathematical Sciences Institute, Cornell University.

Aucher, G., Balbiani, P., del Cerro, L. F., and Herzig, A. (2009). Global and local graph modifiers. *Electronic Notes in Theoretical Computer Science*, 231:293–307.

Aucher, G., van Benthem, J., and Grossi, D. (2015). Sabotage modal logic: Some model and proof theoretic aspects. In van der Hoek, W., Holliday, W., and Wang, W., editors, *Proceedings of LORI '2015*, volume 9394 of *Lecture Notes in Computer Science*, pages 1–13.

Aucher, G., van Benthem, J., and Grossi, D. (2018). Modal logics of sabotage revisited. *Journal of Logic and Computation*, 28:269–303.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books, New York.

Baltag, A., Bezhanishvili, N., Özgün, A., and Smets, S. (2019a). A topological approach to full belief.

Journal of Philosophical Logic, 48:205–244.

Baltag, A., Boddy, R., and Smets, S. (2018). Group knowledge in interrogative epistemology. In van Ditmarsch, H., Sandu, G., and Hintikka, J., editors, *Outstanding Contributions to Logic*, volume 12, pages 131–164. Springer.

Baltag, A., Christoff, Z., Rendsvig, R. K., and Smets, S. (2019b). Dynamic epistemic logics of diffusion and prediction in social networks. *Studia Logica*, 107:489–531.

Baltag, A. and Cinà, G. (2018). Bisimulation for conditional modalities. *Studia Logica*, 106:1–33.

Baltag, A., Li, D., and Pedersen, M. Y. (2019c). On the right path: A modal logic for supervised learning. In Blackburn, P., Lorini, E., and Guo, M., editors, *Proceedings of LORI '2019*, volume 11813 of *LNCS*, pages 1–14. Springer.

Baltag, A., Li, D., and Pedersen, M. Y. (2021). A modal logic for supervised learning. *Journal of Logic, Language and Information (Accepted)*.

Baltag, A., Moss, L., and Solecki, S. (1998). The logic of public announcements and common knowledge and private suspicions. In Gilboa, I., editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK '1998)*, pages 43–56, USA. Morgan Kaufmann.

Baltag, A. and Smets, S. (2020). Learning what others know. In Albert, E. and Kovacs, L., editors, *LPAR-23: 23rd International Conference on Logic for Programming, Artificial Intelligence and Reasoning*, volume 73 of *EPiC Series in Computing*, pages 90–119. EasyChair.

Baltag, A. and van Benthem, J. (2021a). The logic of continuous dependence. Manuscript.

Baltag, A. and van Benthem, J. (2021b). A simple logic of functional dependence. *Journal of Philosophical Logic*.

Belardinelli, F., van Ditmarsch, H., and van der Hoek, W. (2017). A logic for global and local announcements. In Lang, J., editor, *Proceedings of the Sixteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK '2017)*, pages 28–42.

Berto, F. and Tagliabue, J. (2021). Cellular automata. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2021 edition.

Blackburn, P., de Rijke, M., and Venema, Y. (2001). *Modal Logic*. Cambridge University Press, Cambridge, UK.

Blackburn, P. and Seligman, J. (1995). Hybrid languages. *Journal of Logic, Language and Information*, 4:251–272.

Blando, F. Z., Mierzewski, K., and Areces, C. (2020). The modal logics of the poison game. In Liu, F., Ono, H., and Yu, J., editors, *Knowledge, Proof and Dynamics*, Logic in Asia: Studia Logica Library, pages 3–23. Springer.

C. Areces and R. Fervari and G. Hoffmann (2014). Swap logic. *Logic Journal of the IGPL*, 22:309–332.

Carrington, R. M. (2013). Learning and Knowledge in Social Networks. Master's thesis, Institute for Logic, Language and Computation, University of Amsterdam.

Chang, C. C. and Keisler, H. J. (1973). *Model Theory*. Studies in Logic and the Foundations of Mathematics. Elsevier, North-Holland.

Christoff, Z. (2016). *Dynamic Logics of Networks: Information Flow and the Spread of Opinion*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam.

Christoff, Z. and Grossi, D. (2017). Stability in binary opinion diffusion. In Baltag, A., Seligman, J., and Yamada, T., editors, *Proceedings of LORI '2017*, volume 10455 of *LNCS*, pages 166–180. Springer.

Christoff, Z. and Hansen, J. U. (2013). A two-tiered formalization of social influence. In Grossi, D., Roy, O., and Huang, H., editors, *Proceedings of LORI '2013*, volume 8196 of *LNCS*, pages 68–81. Springer.

Christoff, Z. and Hansen, J. U. (2015). A logic for diffusion in social networks. *Journal of Applied Logic*, 13:48–77.

Christoff, Z., Hansen, J. U., and Proietti, C. (2016). Reflecting on social influence in networks. *Journal of Logic, Language and Information*, 25:299–333.

Dégremont, C. and Gierasimczuk, N. (2011). Finite identification from the viewpoint of epistemic update. *Information and Computation*, 209:383–396.

Demey, L. (2011). Some remarks on the model theory of epistemic plausibility models. *Journal of Applied Non-Classical Logics*, 21:375–395.

Easley, D. and Kleinberg, J. (2010). *Networks, Crowds, and Markets*. Cambridge University Press, Cambridge.

Fagin, R., Halpern, J., Moses, Y., and Vardi, M. (1995). *Reasoning about Knowledge*. MIT Press, Cambridge.

Fervari, R. (2014). *Relation-Changing Modal Logics*. PhD thesis, Facultad de Matemática, Universidad Nacional de Córdoba.

Franklin, B. (1786). The morals of chess. *The Columbian Magazine*.

Gabbay, D. (2008). Introducing reactive kripke semantics and arc accessibility. In Avron, A., Dershowitz, N., and Rabinovich, A., editors, *Pillars of Computer Science: Essays Dedicated to Boris (Boaz) Trakhtenbrot on the Occasion of His 85th Birthday*, volume 4800 of *Lecture Notes in Computer Science*, pages 292–341. Springer Verlag.

Gabbay, D. (2013). *Reactive Kripke Semantics*. Springer-Verlag, Berlin.

Gargov, G. and Passy, S. (1990). A note on boolean modal logic. In Petkov, P., editor, *Mathematical Logic*, page 299–309. Springer.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23(6):121–123.

Gierasimczuk, N. (2009a). Bridging learning theory and dynamic epistemic logic. *Synthese*, 169:371–384.

Gierasimczuk, N. (2009b). Learning by erasing in dynamic epistemic logic. In Dediu, A. H., Ionescu, A. M., and Martín-Vide, C., editors, *Proceedings of LATA '09*, volume 5457 of *LNCS*, pages 362–373. Springer.

Gierasimczuk, N. (2010). *Knowing One's Limits: Logical Analysis of Inductive Inference*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam.

Gierasimczuk, N. and de Jongh, D. (2013). On the complexity of conclusive update. *The Computer Journal*, 56:365–377.

Gierasimczuk, N., Kurzen, L., and Velázquez-Quesada, F. (2009). Learning and teaching as a game: A sabotage approach. In He, X., Horty, J., and Pacuit, E., editors, *Proceedings of LORI '2009*, volume 5834 of *Lecture Notes in Computer Science*, pages 119–132.

Goranko, V. and Passy, S. (1992). Using the universal modality: Gains and questions. *Journal of Logic and Computation*, 2:5–30.

Halpern, J. (2016). *Actual Causality*. MIT Press, Cambridge.

Hansen, J. U. (2011). A hybrid public announcement logic with distributed knowledge. *Electronic Notes in Theoretical Computer Science*, 273:33–50.

Harel, D. (1985). Recurring dominoes: Making the highly undecidable highly understandable. In Karplnski, M. and van Leeuwen, J., editors, *Topics in the Theory of Computation*, volume 102 of *North-Holland Mathematics Studies*, pages 51–71. North-Holland.

Hendricks, V. F. and Hansen, P. G. (2016). *Infostorms*. Copernicus.

Hintikka, J. (1973). *Logic, Language-Games and Information*. Oxford University Press.

Hintikka, J. (1998). *The Principles of Mathematics Revisited*. Cambridge University Press, Cambridge.

Hintikka, J. and Sandu, G. (1997). Game-theoretical semantics. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 361–410. Elsevier.

Hodges, W. (1997). Compositional semantics for a language of imperfect information. *Logic Journal of the IGPL*, 5:539–563.

Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge.

Hoffmann, G. (2015). Undecidability of a very simple modal logic with binding. *CoRR*, abs/1508.03630.

Holliday, W. H., Hoshi, T., and Icard, T. (2013). Information dynamics and uniform substitution. *Synthese*, 190:31–55.

Hornischer, L. (2021). *Computation: Symbolic or Non-Symbolic? Toward a Unified Foundation*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam.

Huizinga, J. (1949). *Homo Ludens: A Study of the Play-Element in Culture*. Routledge.

Ibeling, D. and Icard, T. (2020). Probabilistic reasoning across the causal hierarchy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10170–10177.

Johansen, L. (1982). On the status of the nash type of noncooperative equilibrium in economic theory. *The Scandinavian Journal of Economics*, 84:421–441.

Keiff, L. (2011). Dialogical logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2011 edition.

Kelly, K. (1996). *The Logic of Reliable Inquiry*. Oxford University Press, Oxford, UK.

Kelly, K. T., Schulte, O., and Juhl, C. (1997). Learning theory and the philosophy of science. *Philosophy of Science*, 64(2):245–267.

Klein, D. and Rendsvig, R. K. (2017). Convergence, continuity and recurrence in dynamic epistemic logic. In Baltag, A., Seligman, J., and Yamada, T., editors, *Proceedings of LORI '2017*, volume 10455 of *Lecture Notes in Computer Science*, pages 108–122.

Kozen, D. (1983). Results on the propositional μ-calculus. *Theoretical Computer Science*, 27:333–354.

Kremer, P. and Mints, G. (2007). Dynamic topological logic. In Aiello, M., Pratt-Hartmann, I., and van Benthem, J., editors, *Handbook of Spatial Logics*, pages 565–606. Springer.

Lange, S., Zeugmann, T., and Kapur, S. (1996). Monotonic and dual monotonic language learning. *Theoretical Computer Science*, 155:365–410.

Li, D. (2020). Losing connection: The modal logic of definable link deletion. *Journal of Logic and Computation*, 30:715–743.

Li, D. and Wang, Y. (2021). Modal-logical mereology. *Review of Symbolic Logic (Under Review)*.

Liang, Z. and Seligman, J. (2011). A logical model of the dynamics of peer pressure. *Electronic Notes in Theoretical Computer Science*, 278:275–288.

Liu, F., Seligman, J., and Girard, P. (2014). Logical dynamics of belief change in the community. *Synthese*, 191:2403–2431.

Löding, C. and Rohde, P. (2003a). Model checking and satisfiability for sabotage modal logic. In Pandya, P. K. and Radhakrishnan, J., editors, *Foundations of Software Technology and Theoretical Computer Science (FSTTCS '2003)*, volume 2914 of *Lecture Notes in Computer Science*, pages 302–313.

Löding, C. and Rohde, P. (2003b). Solving the sabotage game is PSPACE-hard. In Rovan, B. and Vojtáš, P., editors, *MFCS '2003*, volume 2747 of *Lecture Notes in Computer Science*, pages 531–540.

Mahdi, H. B. (2010). Products of alexandroff spaces. *International Journal of Contemporary Mathematical Sciences*, 5:2037–2047.

Mukouchi, Y. (1992). Characterization of finite identification. In Jantke, K. P., editor, *Analogical and Inductive Inference*, volume 642 of *LNAI*, pages 260–267. Springer.

Osborne, M. and Rubinstein, A. (1994). *A Course in Game Theory*. MIT Press, Cambridge.

Pacuit, E. (2017). *Neighborhood Semantics for Modal Logic*. Springer.

Parikh, R. and Ramanujam, R. (2003). A knowledge based semantics of messages. *Journal of Logic, Language and Information*, 12:453–467.

Paterson, G. (2018). *Speaker Systems*. PhD thesis, Department of Philosophy, Stanford University.

Pedersen, T. and Slavkovik, M. (2017). Formal models of conflicting social influence. In An, B., Bazzan, A. L. C., Leite, J., Villata, S., and van der Torre, L. W. N., editors, *PRIMA '2017*, volume 10621 of *LNCS*, pages 349–365. Springer.

Plaza, J. A. (1989). Logics of public communications. In Emrich, M. L., Pfeifer, M. S., Hadzikadic, M., and Ras, Z. W., editors, *Proceedings of the 4th international symposium on methodologies for intelligent systems*, pages 201–216. Oak Ridge National Laboratory.

Rendsvig, R. K. (2014). Pluralistic ignorance in the bystander effect: Informational dynamics of unresponsive witnesses in situations calling for intervention. *Synthese*, 191:2471–2498.

Rohde, P. (2005). *On Games and Logics over Dynamically Changing Structures*. PhD thesis, Fakultät für Mathematik, Informatik und Naturwissenschaften, RWTH Aachen University.

Seligman, J., Liu, F., and Girard, P. (2011). Logic in the community. In Banerjee, M. and Seth, A., editors, *Logic and Its Applications (ICLA '2011)*, volume 6521 of *LNCS*, pages 178–188. Springer.

Shi, C. (2021). Collective opinion as tendency towards consensus. *Journal of Philosophical Logic*, 50:593–613.

Shi, C. and Wang, Y. (2021). Pareto optimality, functional dependence and collective agent. Manuscript, Joint Research Center for Logic, Tsinghua University.

Shoham, Y. and Leyton-Brown, K. (2008). *Multiagent Systems: Algorithmic, Game-Theoretic and Logical Foundations*. Cambridge University Press, Cambridge.

Skyrms, B. (1990). *The Dynamics of Rational Deliberation*. Harvard University Press, Cambridge.

Smets, S. and Velázquez-Quesada, F. R. (2017). How to make friends: A logical approach to social group creation. In Baltag, A., Seligman, J., and Yamada, T., editors, *Proceedings of LORI '2017*, volume 10455 of *LNCS*, pages 377–390. Springer.

Smets, S. and Velázquez-Quesada, F. R. (2020). A closeness- and priority-based logical study of social network creation. *Journal of Logic, Language and Information*, 29:21–51.

ten Cate, B. and Franceschet, M. (2005). On the complexity of hybrid logics with binders. In Ong, L., editor, *Proceedings of Computer Science Logic 2005*, volume 3634 of *Lecture Notes in Computer Science*, pages 339–354.

Thompson, D. (2020). Local fact change logic. In Liu, F., Ono, H., and Yu, J., editors, *Knowledge, Proof and Dynamics*, Logic in Asia: Studia Logica Library, pages 73–96. Springer.

Väänänen, J. (2007). *Dependence Logic: A New Approach to Independence Friendly Logic*. Cambridge University Press, Cambridge.

van Benthem, J. (1984). Correspondence theory. In Gabbay, D. and Guenthner, F., editors, *Handbook of Philosophical Logic: Volume II: Extensions of Classical Logic*, pages 167–247. Springer.

van Benthem, J. (1996). *Exploring Logical Dynamics*. CSLI Publication, California.

van Benthem, J. (2005). An essay on sabotage and obstruction. In Hutter, D. and Stephan, W., editors, *Mechanizing Mathematical Reasoning*, volume 2605 of *LNCS*, pages 268–276. Springer.

van Benthem, J. (2006). "One is a lonely number": Logic and communication. In Chatzidakis, Z., Koepke, P., and Pohlers, W., editors, *Logic Colloquium '02*, Lecture Notes in Logic, page 96–129. Cambridge University Press.

van Benthem, J. (2010). *Modal Logic for Open Minds*. CSLI Publications, California.

van Benthem, J. (2011). *Logical Dynamics of Information and Interaction*. Cambridge University Press, Cambridge, UK.

van Benthem, J. (2014). *Logic in Games*. MIT Press, Cambridge.

van Benthem, J. (2015). Oscillations, logic, and dynamical systems. In Ghosh, S. and Szymanik, J., editors, *The Facts Matter*, pages 9–22. College Publications.

van Benthem, J. and Bezhanishvili, G. (2007). Modal logics of space. In Aiello, M., Pratt-Hartmann, I., and van Benthem, J., editors, *Handbook of Spatial Logics*, pages 217–298. Springer.

van Benthem, J., Bezhanishvili, N., and Enqvist, S. (2019). A new game equivalence, its logic and algebra. *Journal of Philosophical Logic*, 48:649–684.

van Benthem, J. and Klein, D. (2020). Logics for analyzing games. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2020 edition.

van Benthem, J., Li, L., Shi, C., and Yin, H. (2021a). Hybrid sabotage modal logic. Manuscript, Joint Research Center for Logic, Tsinghua University.

van Benthem, J. and Liu, F. (2007). Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17:157–182.

van Benthem, J. and Liu, F. (2020). Graph games and logic design. In Liu, F., Ono, H., and Yu, J., editors, *Knowledge, Proof and Dynamics*, Logic in Asia: Studia Logica Library, pages 125–146. Springer.

van Benthem, J., Liu, F., and Smets, S. (2021b). Logico-computational aspects of rationality. In Knauff, M. and Spohn, W., editors, *Handbook of Rationality (to appear)*. The MIT Press.

van Benthem, J., Mierzewski, K., and Blando, F. Z. (2020). The modal logic of stepwise removal. *The Review of Symbolic Logic*.

van Benthem, J., van Eijck, J., and Kooi, B. (2006). Logics of communication and change. *Information and Computation*, 204:1620–1662.

van Ditmarsch, H., van der Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*, volume 337 of *Synthese Library*. Springer.

Varzi, A. (2019). Mereology. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2019 edition.

Wang, Y., Sietsma, F., and van Eijck, J. (2010). Logic of information flow on communication channels. In Omicini, A., Sardina, S., and Vasconcelos, W., editors, *DALT '2010*, volume 6619 of *LNCS*, pages 130–147. Springer.

Wooldridge, M. (2002). *An Introduction to MultiAgent Systems*. John Wiley, New York.

Xie, K. (2020). *Where Causality, Conditionals and Epistemology Meet: A Logical Inquiry*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam.

Zhang, T. (2020). Solution complexity of local variants of sabotage game. In Liu, F., Ono, H., and Yu, J., editors, *Knowledge, Proof and Dynamics*, Logic in Asia: Studia Logica Library, pages 97–121. Springer.

Acknowledgements

First of all, my most important word of thanks goes to my supervisors at Tsinghua and ILLC: Fenrong Liu, Johan van Benthem and Alexandru Baltag, who together created the most supportive group I can imagine. In the last four years, they gave me enough freedom to pursue my own research interests as well as much guidance to keep realistic goals and strict timelines. I want to express my great gratitude to Fenrong for providing me with the opportunity to be a PhD student and giving me the constant and comprehensive support throughout all stages of my studies at the two institutes. I am deeply indebted to Johan, who gave me a great deal of help in the last four years and taught me a lot on logic and many other aspects including academic writing and even English grammar. Also, I owe a great debt of thanks to Alexandru, who gave me the chance to study at the ILLC and spent plenty of time and patience in guiding my research. I would like thank all of them for their guidance, support and patience during these years, which made my studies colorful. I was lucky to have all of them as my supervisors.

I thank Franz Berto, Martin Stokhof, Hao Tang, Yde Venema, Wei Wang and Donghua Zhu for agreeing to be in my thesis committee.

I am grateful to all my co-authors of the papers contained in the thesis: Alexandru Baltag, Johan van Benthem, Mina Young Pedersen, and Fernando Raymundo Velázquez Quesada. A special thank you goes to Yanjing Wang: although the work with Yanjing is not included here, what I gained from the collaboration and his guidance also nourished the thesis indirectly.

I would like to thank other professors I met in Amsterdam, Tsinghua or online, Nick Bezhanishvili, Peter van Emde Boas, Sujata Ghosh, Davide Grossi, Jeremy Seligman, Sonja Smets, Martin Stokhof, Lu Wang, Dag Westerståhl, and Junhua Yu. I benefited a lot from their courses or the discussions with them.

Also, I benefited a lot from discussions with others, including Peihui Chen, Yu Chen, Haibin Gui, Mingliang Liu, Dean McHugh, Mina Young Pedersen, Carlo Proietti, Chenwei Shi, Zhiqiang Sun, Fernando Raymundo Velázquez Quesada, Yu Wei, Kaibo Xie, Chao Xu, Antonio Yuste-Ginel and many others.

Finally, the warmest thanks go to my family for all their support in all my decisions, which gave me the freedom and courage to pursue what I am interested in.

Résumé and Academic Achievements

Résumé

Dazhu Li was born on the 17th of November 1991 in Jiaozuo, Henan, China.

He began his bachelor's study in the Department of Philosophy, Zhengzhou University in September 2009, majoring in philosophy, and got a Bachelor of Philosophy degree in July 2013.

He began his master's study in the Department of Philosophy, Peking University in September 2013, and got a Master of Philosophy degree in Logic in July 2017.

In September 2017, he started to pursue a doctorate in the Department of Philosophy, Tsinghua University. In March 2018, he was admitted to the jointly awarded doctorate program of Tsinghua University and the Institute for Logic, Language and Computation, University of Amsterdam.

Academic Achievements

[1] Alexandru Baltag, Dazhu Li, Mina Young Pedersen. On the right path: a modal logic for supervised learning. *Proceedings of LORI 2019*, *Lecture Notes in Computer Science*, 2019, 11813:1-14.

[2] Dazhu Li. Losing connection: the modal logic of definable link deletion. *Journal of Logic and Computation*, 2020, 30(3):715-743.

[3] Alexandru Baltag, Dazhu Li, Mina Young Pedersen. A modal logic for supervised learning. *Journal of Logic, Language and Information*. Accepted. 2021.

[4] Dazhu Li, Sujata Ghosh, Fenrong Liu, Yaxin Tu. On the subtle nature of a simple logic of the hide and seek game. To be presented at WoLLIC 2021. 2021.

[5] Dazhu Li, Yanjing Wang. Modal-logical mereology. *Review of Symbolic Logic*. Under review. 2021.

[6] Alexandru Baltag, Johan van Benthem, Dazhu Li. Logical proposals for dynamic dependence. Draft. 2021.

Titles in the ILLC Dissertation Series:

ILLC DS-2016-01: **Ivano A. Ciardelli**
Questions in Logic

ILLC DS-2016-02: **Zoé Christoff**
Dynamic Logics of Networks: Information Flow and the Spread of Opinion

ILLC DS-2016-03: **Fleur Leonie Bouwer**
What do we need to hear a beat? The influence of attention, musical abilities, and accents on the perception of metrical rhythm

ILLC DS-2016-04: **Johannes Marti**
Interpreting Linguistic Behavior with Possible World Models

ILLC DS-2016-05: **Phong Lê**
Learning Vector Representations for Sentences - The Recursive Deep Learning Approach

ILLC DS-2016-06: **Gideon Maillette de Buy Wenniger**
Aligning the Foundations of Hierarchical Statistical Machine Translation

ILLC DS-2016-07: **Andreas van Cranenburgh**
Rich Statistical Parsing and Literary Language

ILLC DS-2016-08: **Florian Speelman**
Position-based Quantum Cryptography and Catalytic Computation

ILLC DS-2016-09: **Teresa Piovesan**
Quantum entanglement: insights via graph parameters and conic optimization

ILLC DS-2016-10: **Paula Henk**
Nonstandard Provability for Peano Arithmetic. A Modal Perspective

ILLC DS-2017-01: **Paolo Galeazzi**
Play Without Regret

ILLC DS-2017-02: **Riccardo Pinosio**
The Logic of Kant's Temporal Continuum

ILLC DS-2017-03: **Matthijs Westera**
Exhaustivity and intonation: a unified theory

ILLC DS-2017-04: **Giovanni Cinà**
Categories for the working modal logician

ILLC DS-2017-05: **Shane Noah Steinert-Threlkeld**
Communication and Computation: New Questions About Compositionality

ILLC DS-2017-06: **Peter Hawke**
The Problem of Epistemic Relevance

ILLC DS-2017-07: **Aybüke Özgün**
Evidence in Epistemic Logic: A Topological Perspective

ILLC DS-2017-08: **Raquel Garrido Alhama**
Computational Modelling of Artificial Language Learning: Retention, Recognition & Recurrence

ILLC DS-2017-09: **Miloš Stanojević**
Permutation Forests for Modeling Word Order in Machine Translation

ILLC DS-2018-01: **Berit Janssen**
Retained or Lost in Transmission? Analyzing and Predicting Stability in Dutch Folk Songs

ILLC DS-2018-02: **Hugo Huurdeman**
Supporting the Complex Dynamics of the Information Seeking Process

ILLC DS-2018-03: **Corina Koolen**
Reading beyond the female: The relationship between perception of author gender and literary quality

ILLC DS-2018-04: **Jelle Bruineberg**
Anticipating Affordances: Intentionality in self-organizing brain-body-environment systems

ILLC DS-2018-05: **Joachim Daiber**
Typologically Robust Statistical Machine Translation: Understanding and Exploiting Differences and Similarities Between Languages in Machine Translation

ILLC DS-2018-06: **Thomas Brochhagen**
Signaling under Uncertainty

ILLC DS-2018-07: **Julian Schlöder**
Assertion and Rejection

ILLC DS-2018-08: **Srinivasan Arunachalam**
Quantum Algorithms and Learning Theory

ILLC DS-2018-09: **Hugo de Holanda Cunha Nobrega**
Games for functions: Baire classes, Weihrauch degrees, transfinite computations, and ranks

ILLC DS-2018-10: **Chenwei Shi**
Reason to Believe

ILLC DS-2018-11: **Malvin Gattinger**
New Directions in Model Checking Dynamic Epistemic Logic

ILLC DS-2018-12: **Julia Ilin**
Filtration Revisited: Lattices of Stable Non-Classical Logics

ILLC DS-2018-13: **Jeroen Zuiddam**
Algebraic complexity, asymptotic spectra and entanglement polytopes

ILLC DS-2019-01: **Carlos Vaquero**
What Makes A Performer Unique? Idiosyncrasies and commonalities in expressive music performance

ILLC DS-2019-02: **Jort Bergfeld**
Quantum logics for expressing and proving the correctness of quantum programs

ILLC DS-2019-03: **András Gilyén**
Quantum Singular Value Transformation & Its Algorithmic Applications

ILLC DS-2019-04: **Lorenzo Galeotti**
The theory of the generalised real numbers and other topics in logic

ILLC DS-2019-05: **Nadine Theiler**
Taking a unified perspective: Resolutions and highlighting in the semantics of attitudes and particles

ILLC DS-2019-06: **Peter T.S. van der Gulik**
Considerations in Evolutionary Biochemistry

ILLC DS-2019-07: **Frederik Möllerström Lauridsen**
Cuts and Completions: Algebraic aspects of structural proof theory

ILLC DS-2020-01: **Mostafa Dehghani**
Learning with Imperfect Supervision for Language Understanding

ILLC DS-2020-02: **Koen Groenland**
Quantum protocols for few-qubit devices

ILLC DS-2020-03: **Jouke Witteveen**
Parameterized Analysis of Complexity

ILLC DS-2020-04: **Joran van Apeldoorn**
A Quantum View on Convex Optimization

ILLC DS-2020-05: **Tom Bannink**
Quantum and stochastic processes

ILLC DS-2020-06: **Dieuwke Hupkes**
Hierarchy and interpretability in neural models of language processing

ILLC DS-2020-07: **Ana Lucia Vargas Sandoval**
On the Path to the Truth: Logical & Computational Aspects of Learning

ILLC DS-2020-08: **Philip Schulz**
Latent Variable Models for Machine Translation and How to Learn Them

ILLC DS-2020-09: **Jasmijn Bastings**
A Tale of Two Sequences: Interpretable and Linguistically-Informed Deep Learning for Natural Language Processing

ILLC DS-2020-10: **Arnold Kochari**
Perceiving and communicating magnitudes: Behavioral and electrophysiological studies

ILLC DS-2020-11: **Marco Del Tredici**
Linguistic Variation in Online Communities: A Computational Perspective

ILLC DS-2020-12: **Bastiaan van der Weij**
Experienced listeners: Modeling the influence of long-term musical exposure on rhythm perception

ILLC DS-2020-13: **Thom van Gessel**
Questions in Context

ILLC DS-2020-14: **Gianluca Grilletti**
Questions & Quantification: A study of first order inquisitive logic

ILLC DS-2020-15: **Tom Schoonen**
Tales of Similarity and Imagination. A modest epistemology of possibility

ILLC DS-2020-16: **Ilaria Canavotto**
Where Responsibility Takes You: Logics of Agency, Counterfactuals and Norms

ILLC DS-2020-17: **Francesca Zaffora Blando**
Patterns and Probabilities: A Study in Algorithmic Randomness and Computable Learning

ILLC DS-2021-01: **Yfke Dulek**
Delegated and Distributed Quantum Computation

ILLC DS-2021-02: **Elbert J. Booij**
The Things Before Us: On What it Is to Be an Object

ILLC DS-2021-03: **Seyyed Hadi Hashemi**
Modeling Users Interacting with Smart Devices

ILLC DS-2021-04: **Sophie Arnoult**
Adjunction in Hierarchical Phrase-Based Translation

ILLC DS-2021-05: **Cian Guilfoyle Chartier**
A Pragmatic Defense of Logical Pluralism

ILLC DS-2021-06: **Zoi Terzopoulou**
Collective Decisions with Incomplete Individual Opinions

ILLC DS-2021-07: **Anthia Solaki**
Logical Models for Bounded Reasoners

ILLC DS-2021-08: **Michael Sejr Schlichtkrull**
Incorporating Structure into Neural Models for Language Processing

ILLC DS-2021-09: **Taichi Uemura**
Abstract and Concrete Type Theories

ILLC DS-2021-10: **Levin Hornischer**
Dynamical Systems via Domains: Toward a Unified Foundation of Symbolic and Non-symbolic Computation

ILLC DS-2021-11: **Sirin Botan**
Strategyproof Social Choice for Restricted Domains

ILLC DS-2021-12: **Michael Cohen**
Dynamic Introspection

ILLC DS-2021-13: **Dazhu Li**
Formal Threads in the Social Fabric: Studies in the Logical Dynamics of Multi-Agent Interaction

ILLC DS-2022-01: **Anna Bellomo**
Sums, Numbers and Infinity: Collections in Bolzano's Mathematics and Philosophy

ILLC DS-2022-02: **Jan Czajkowski**
Post-Quantum Security of Hash Functions

ILLC DS-2022-03: **Sonia Ramotowska**
Quantifying quantifier representations: Experimental studies, computational modeling, and individual differences