Dynamic logics of networks: Information flow and the spread of opinion

Christoff, Z.L.

Publication date
2016
Document Version
Final published version

Institute for Logic, Language and Computation

Dynamic Logics of Networks

Zoé Christoff

Dynamic Logics of Networks

Information Flow and the Spread of Opinion

Zoé Christoff

Dynamic Logics of Networks

Information Flow and the Spread of Opinion

Zoé Christoff

Dynamic Logics of Networks

Information Flow and the Spread of Opinion

ILLC Dissertation Series DS-2016-02

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/

The investigations for this dissertation were supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no.283963.

Cover design by Nina Gierasimczuk.
Printed and bound by Ipskamp Drukkers.

ISBN: 978-94-028-0081-4

Dynamic Logics of Networks

Information Flow and the Spread of Opinion

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het College voor
Promoties ingestelde commissie, in het openbaar
te verdedigen in de Agnietenkapel
op dinsdag 22 maart 2016, te 10.00 uur

door

Zoé Laure Christoff

geboren te Genève, Zwitserland.

Promotor:	Prof. dr. J.F.A.K. van Benthem	Universiteit van Amsterdam
Co-promotor:	Dr. S.J.L. Smets	Universiteit van Amsterdam
Overige leden:	Prof. dr. J. van Eijck	Universiteit van Amsterdam
	Dr. D. Grossi	Universiteit van Liverpool
	Prof. dr. V.F. Hendricks	Universiteit van Kopenhagen
	Prof. dr. F. Liu	Universiteit van Amsterdam
	Prof. dr. O. Roy	Universiteit van Bayreuth
	Prof. dr. F.J.M.M. Veltman	Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Contents

Acknowledgments

I am grateful to the many people who have contributed to this thesis by their support, their trust, their collaboration, their encouragement, or their advice.

First of all, I would like to thank both my supervisors, Johan van Benthem and Sonja Smets. Together, they provided me with a combination of enough freedom to pursue my own ideas and projects, yet enough structure to keep realistic goals and a strict timeline.

I am grateful to Sonja who gave me the opportunity to join her project "The Logic of Correlated Information Change" almost four years ago. I have felt lucky ever since, not only to be part of a wonderful team, but also to work on exactly the topics I was enthusiastic about. Sonja encouraged me from the start to be as active as possible: she pushed me to submit papers to workshops and conferences, to go to summer schools, to visit people abroad, and to give many talks. I have immensely benefited from her advice and her trust, both academically and humanly. On a more personal note, I would also like to thank Sonja for being such a caring person during those four years.

I thank Johan for his invaluable supervision and guidance. We have had many meetings, filled in the blackboard a few times, made many plans, made many new plans, and touched upon an incredible variety of topics in our discussions. Not only did Johan give me precious feedback on my work, always keeping "the big picture" in mind while paying attention to the detail, he also made me believe that I would succeed in writing this book. He helped me more than I can describe here and I am extremely grateful for his time, his advice, his honesty, and his encouragement.

I thank Jan van Eijck, Davide Grossi, Vincent F. Hendricks, Fenrong Liu, Olivier Roy, and Frank Veltman for accepting to be members of my thesis committee.

I am directly indebted to the co-authors of the papers included in this thesis, Alexandru Baltag, Jens Ulrik Hansen, Carlo Proietti, Rasmus K. Rendsvig, and Sonja, as well as to the collaborators on ongoing projects which nourished the thesis more indirectly: Nina Gierasimczuk, and Johan. I owe a special thank you to Jens for being my sparring partner from the start with respect to many of the topics discussed in this thesis, for numerous fruitful discussions in Amsterdam and a few extra ones in Copenhagen, for serious times and for fun times, as well as for being my friend ever since. I owe another special thank you to Nina, who has been at the same time a helpful advisor during those four years and a true friend. Thank you for being a constant when everything else was changing. And thank you for designing the cover of this book.

This thesis also benefited from input from numerous other researchers. While the list is too long to write here, I am particularly grateful to Fenrong Liu and Jeremy Seligman. In addition to our discussions on logics for social networks, I thank Fenrong for giving me the chance to visit her in Beijing, and Jeremy for being a wonderful guide in Chongqing, in Beijing, and in Hangzhou, from street barbecues and tea shops to mahjong and karaoke. Those are unforgettable times.

I am also grateful to the ILLC for providing the best possible environment for a PhD, and especially to Jenny Batson, Karine Gigengack, and Tanja Kassenaar for helping me in all practical matters of life as a PhD candidate in Amsterdam.

My time in Amsterdam would have been very different without the friends I have met through the ILLC: Aybüke, Ben, Dora, Facundo, Jakub, Jens, Julia, Max, Nick, Nina, Paula, Soroush, Virginie, Inés, and outside the ILLC: Amber, Bas, Camille, Charley, Liselotte, Marisa, Sarah, Smriti, and Marc. Thank you for making Amsterdam feel like home.

I take this opportunity to thank the teachers who inspired me and encouraged me to embark on this PhD journey: Alexandru Baltag, in the ILLC, as well as Kevin Mulligan, Fabrice Correia, Pascal Engel, and Philipp Blum, in the Department of Philosophy of the University of Geneva. Among earlier influences, I will mention only one: my high-school teacher of applied mathematics, Michel Küne, without whom I might have never discovered that I liked graph theory, tiling problems, and philosophy. This thesis would not exist without them.

Last but not least, thanks to my "Swiss family" for never failing to support me: Annika, Carole, Clémence, Crystel, Emilda, Julide, Marie, Mary, Morganne, Sandie, and Thao.

Zoé Christoff

Amsterdam, February 2016.

Part I

Preliminaries

Chapter 1

Introduction

1.1 Information and Flow of Opinion in Social Networks

Information and reasoning in social networks. Humans are inherently social beings that constantly influence each other. We all live in a web of social networks that shape our opinions and behavior.

The importance of networks has long been recognized in various disciplines, using different methodologies. The study of social influence, conceived as the way "other people affect one's beliefs, feelings and behavior" [122], lies at the heart of social psychology, where much research takes the form of controlled laboratory experiments with actual people.

Placing the focus on a higher aggregation level, scholars from other traditions, such as economics, sociology, and political science, have studied social influence via large-scale abstract social networks models (see [105], [71]). Typically, such network models abstract away from much of the micro-level complexity of individual psychological processes. Nevertheless, they are ideal for modeling the effect of social influence over time and over different configurations of social groups, including the emergence of broader high-level patterns of group behavior. Two characteristic examples, out of many, are models for diffusion of innovations [90] or of creation of micro-cultures [14].[1]

In addition to these more empirical approaches, techniques for modeling and designing networks and group behavior are also a prominent theme in mathematics [164], computer science [101, 7, 11, 9, 73, 151, 72, 161, 150], and in philosophy [155]. In particular, modern computational systems are societies of agents that have been studied extensively [149, 165] and one important theme in this study is the fact that agents do not just form sets of separate individuals, but intercon-

[1] Networks are also crucial in cognitive neuroscience [156], and 'societies of interacting neurons' sometimes show striking resemblances with models of social behavior.

nected networks where social relations of neighborhood, informational access, and hierarchy play a crucial role[2]. Indeed, theory and societal practice interact here, since new networks are being created all the time, such as Facebook, Twitter, and the like, with sometimes unintended and disturbing emergent behavior such as informational bubbles and other public opinion phenomena that seem at odds with the deliberation presupposed by a democratic society with rational citizens [100]. It seems fair to say that theory is often running far behind these new phenomena, and that we may not yet have the proper conceptual apparatus to understand our situation, let alone, improve it. For an interesting and pioneering attempt at changing this situation, see the analysis of various information-driven social phenomena, and the plan of action outlined in [99, 100].

Against this background, here is what this thesis is about. We believe that at the heart of social behavior, in both its better and worse manifestations, there lies information flow and reasoning. And these are precisely the core topics of the discipline of *logic* – and so, even though logicians can definitely not claim exclusive insights into social structure, given the wide range of other disciplines involved, adding perspectives from logic may be of use.

In particular, in this thesis, we will investigate how social influence and information flow are entangled, and what laws of logic govern this interaction. This will provide a logical perspective on the dynamics of social phenomena over networks, ranging from local social interactions to long-term group behavior: how agents influence each other, how behavior spreads within a population, how such diffusion interacts with information flow, and how the resulting dynamical processes evolve in the limit.

Three major themes in social behavior. Before we explain the contents of this thesis in more detail, we mention three main themes motivating and guiding our research throughout, and thus unifying the different pieces of work presented here. Together, they represent what is striking and challenging about social interaction as we see it.

The first recurrent theme is what we see as the balance between *"transparency"*, the idea that agents can "see" their own minds, and *"opacity"*, the partially hidden nature of other agents' minds: Agents are typically considered to be mostly transparent to themselves, having privileged access to their own mental states, knowing better than anyone else how they feel, what their opinions are, and so on. But agents are much more "opaque" to each other, lacking direct introspection of other minds. However, they may get to learn about each other's mental state by communicating with each other, or by interpreting observed behavior. But this is only partial information, and as such, an agent's interpretation of others may be wrong.

The formal logical tools of this thesis allow us to analyze this balance as it occurs in social phenomena. In particular, we will observe and explain how it is

[2]Also, in the realm of distributed computing, such interrelations are crucial, see e.g. [74].

often the "semi-opacity" of agents which causes surprising results at the group level. In the first part of the thesis, we show how counterintuitive situations of collective failures, such as informational cascades and pluralistic ignorance, are the collective result of rational agents with full reasoning powers that are able to observe each other's actions but not each other's reasons for these actions. In this process, we will analyze what sort of information processing and reasoning can guide agents of different types.

The second major theme, emerging from the first, is the transition from *microbehavior* of individual agents to *macrobehavior* of groups and systems that may show emergent patterns. While the preceding discussion emphasized discrepancies between rationality and achieving desirable goals at the two levels, there are many more things to be understood here. In particular, we aim to capture significant logical laws of collective behavior, connecting the individual level of agents with knowledge and beliefs that can reason and deliberate and agents in networks that can be treated as essentially "parallel automata" reacting to their social environment, repeatedly.

Hence, this thesis is also a logical investigation of long-term agent behavior, where we are interested in questions such as the following. Will the current process result in "public opinion" for the whole group of agents that stabilizes, or will it oscillate, or even diverge? What happens when we assume diversity of agents, even when we treat them as automata with fixed responses of various kinds? The papers in the second part of the thesis propose just the general toolbox to reason about such diffusion phenomena. We will first propose a general dynamic framework allowing to "plug-in" different types of dynamics corresponding to different types of agents. We then enrich this framework to include the modeling of information and information change. And finally, we will propose a minimal framework to model the diffusion of behavior ruled by a given threshold, when agents adopt a behavior as soon a certain proportion of their neighbors have. While the tools we use are of different types (propositional, modal, or hybrid logics), the unity in our toolbox comes from the use of *dynamic logics*.

A third theme that runs throughout this thesis is the effect of social *network structures*. Agent behavior is not just a matter of individual choices: it is constrained in many ways by a social environment that determines what informational access is possible, what the "relevant others" are, and what actions are available. This network structure plays right through the two preceding themes that we identified, since it is a crucial factor in both individual micro-behavior and global system macro-behavior.

Until recently, most of the literature in logics of agency considered either individual actors, or multi-agent interactions based on bare sets of agents, without an explicit account of group network structure. This is true for logics of knowledge and belief [24, 70], of action [114], and even of most logics that have been used to analyze games [132, 43, 37]. This is not to say that these research traditions lack depth or strength: indeed, we will build squarely on this existing work. But in this thesis, we will enrich the perspective, and throughout, take the effects of group structure on board when analyzing the dynamics of information and the spread of opinion in social groups.

What is to follow. In the sections that follow, we say a bit more about the existing work on social networks that forms the direct backdrop to our work, and we briefly place our logic-based approach within the area of logics of agency. After that, we describe the contents of this thesis in more detail, chapter by chapter, and we end by listing the publications on which this thesis is based.

1.2 Our Inspiration: Earlier Approaches to Social Networks

General technical approaches. Our study of the dynamics of social phenomena, such as the spread of behavior or opinions in networks, borrows insights from a number of different fields. Most importantly, the backdrop to our work is the large literature on logics of agency that has started in the 1980s, and still shows no signs of diminishing. This broad research program brings together themes from computer science, computational and philosophical logic, and we cannot even begin to summarize it here. We refer the reader to [149, 165] for up-to-date textbooks. The dynamic turn in logic [32] has also shifted the focus inside logic itself towards the study of informational processes and agent-based interaction.

More specifically, our methodology touches upon work done in philosophy, in particular in the area of social epistemology [167, 84, 85], while also drawing input from philosophical studies of information and the information society [99, 100]. We also take many cues from the large existing body of work on social network theory, among which [127, 71, 105]. Finally, as will be clear from references in specific chapters below, we benefit from applied studies in the social and behavioral sciences, economics, cognitive science, biology, and computer science. While each of the latter areas comes with its own focus, studying specific types of populations, group behavior and dynamic processes, a methodology of formal methods runs across them. These range from purely theoretical models based on a formal language (be it probability theory, logic, or graph theory) to philosophical analysis and the use of experimental data or even simulations with agent-based modeling. Indeed much important network research has been in the tradition of probabilistic and dynamical systems methods [153, 154].

Developments on the logic side. In this thesis, the formal tool that we will mainly use is logic, together with philosophical analysis where appropriate. In doing so, we draw on the insights obtained in modeling individual attitudes (knowledge, beliefs, intentions, preferences, desires, and others) [102, 123, 149, 119, 42, 26, 138], group attitudes (common knowledge, distributed knowledge, common belief, and so on) [76, 158, 36], as well as the dynamics of information change (knowledge updates or belief change) [44, 70, 24, 42, 27]. Our work augments this epistemic/doxastic dimension with a social dimension, by taking the social structure of a group into account. These two dimensions were brought together for the first time only recently.

Let us take a moment to reflect upon why it has taken so long to start building bridges between logic and social network analysis. The past lack of interaction may be explained by their distinct paradigmatic cases of inspiration. On the one side, in social network analysis, one is typically concerned with diffusion phenomena such as the spreading of infections, where agents are taken to be simple bacteria-like automata, reacting all in the same way and all at the same time, to their most direct environment, uniformly, and repeatedly. On the other side, in the dynamic-epistemic logic tradition (and, by and large, in logics of agency as a whole), agents are typically taken to come equipped with unlimited higher-order reasoning powers aiming for the truth, updating their knowledge as they receive new information. Moreover, informational events are typically taken to be sequential. As a result, until a few years ago, the effects of how groups of agents are structured – who communicates with whom, who interacts with whom, who is influenced by whom – have been mostly left aside.

Still, it is fair to say that the idea of structuring groups was in the air: *private* announcements [24] already distinguish between subgroups of "insiders" and "outsiders", [23], where private announcements only affect the insiders, leaving the outsiders clueless about it. By contrast, public announcements reach the whole group uniformly. Likewise, long-term dynamics of iterated update processes had made its entrance earlier on, as witness the long-term limit scenarios of announcements studied in [41], and the work on iterated belief revision policies and convergence to the truth in the long run [29, 30, 81, 20, 21, 22, 108]. A final related earlier research strand is the work on fixed points of belief revision policies corresponding to different attitudes of trust towards information sources [104, 135, 25].

Our most direct sources of inspiration. While acknowledging all the above, the most direct influences on our style of analysis in this thesis are the following two.

First, the work of Girard, Liu, and Seligman makes an inspirational jump towards social structure in their seminal paper "Logic in the Community" [145]. The authors introduce a formal language to talk about networks and the knowledge of agents in these, in a way that meets the received standards of epistemic logic, dynamic logic, while employing notions from hybrid logics to represent indexical aspects of social discourse. They also develop variations of the setting to model preference change ([166], drawing on [119]), as well as belief change [120] under social influence. This line of logics that explicitly deal with social networks has been taken up and refined by various authors [146, 83, 139, 141, 58, 94].

A second slightly different but important direction in logical studies of agency puts the focus on communication networks and temporal protocols for sharing information in groups. Recent work in this line is given by [7, 11, 72] and combines tools from epistemic logic with techniques from distributed computing. This path was paved by earlier work on different communication types, channels, sequences and protocols [16, 136, 31, 161, 73, 151, 150, 69].

The preceding two logical traditions underlie the work in this thesis, which aims to strengthen the connections between logic and social network analysis in further ways.

Further directions: social choice and aggregation. It is also important to stress one more area in which logical methods are gaining importance for studying the social domain. Current work on *social choice theory*, in particular voting theory, judgment aggregation and preference aggregation, has growing connections with recent developments in logic [31, 65, 88]. Different from our emphasis on network structure and communication protocols, in social choice theory the focus often lies on merging information, judgements, preferences or opinions in the absence of individual communication. Such aggregation procedures, too, can of course be taken as a basis for diffusion processes in social networks.

Part of the work we do in this thesis, especially in our second part, on the spread of behavior and opinions makes use of update mechanisms that can also model types of influence coming from aggregation procedures in social choice theory. While we do not make this trajectory explicit, the recent work in [89] illustrates the basic ideas of this connection for the case of opinion diffusion. However, working in a communication-free environment poses some restrictions that are not assumed in this thesis.[3] Clearly, in social reality, agents can be influenced by others via a whole range of methods, including voting, observations, and explicit acts of communication. Accordingly, for many of the models that we design for social scenarios in this thesis, aggregation procedures can form the basis of influence mechanisms, but so can communication and deliberation. We will also discuss the border line between the two in our second part.

1.3 Outline of the Thesis

This thesis uses logical tools to address a number of central issues about social phenomena in networks that have been identified in broad outline in our first

[3]This difference touches upon a debate in the literature on rational consensus formation (using, e.g., weighted models [68, 117]) on how to align aggregation procedures in social choice theory with a perspective of deliberation and communication [56].

section: the interplay of transparent and opaque knowledge about other agents, the transition from micro- to macro-behavior, and the crucial role of network structure in all of this in addition to the capabilities of individual agents. Each part and each chapter provides a window for applying logic to address such issues.

Here is a brief overview of the content. Our parts and chapters also come with brief further introductions and conclusions elaborating our story line.

Part I contains the preliminaries for the presentation of our research. Following the introduction and motivation of our topics in Chapter 1, Chapter 2 introduces some basic technical background material for what follows, coming both from logic and from social networks analysis.

Part II focuses on the following question. When and how does individual rationality lead to group success or group failure? In particular, can individuals who are behaving perfectly rationally lead a group to collective failure? We present logical case studies of two well-known counterintuitive social effects: informational cascades and pluralistic ignorance, where individual rationality can lead to some form of group failure. In the process we will see precisely why this happens, giving insight into how to change individual settings in order to change collective outcomes.

Chapter 3 uses two variants of dynamic-epistemic logic to show how perfectly rational agents who start following the crowd may get stuck in an "informational cascade" leading everybody to be wrong, despite the availability of enough evidence to avoid such a catastrophe. Our formal modeling confirms that, whether agents are full-fledge Bayesian probabilists or use simpler reasoning methods, and whether they have unbounded higher-order reasoning or not, informational cascades are indeed inescapable: even individuals reasoning to the best of their ability sometimes lead their whole community towards an epistemic catastrophe.

Chapter 4 models within hybrid logic a second counterintuitive social phenomenon: "pluralistic ignorance", in which all agents are mistaken about each other's beliefs and all wrongly believe that their own beliefs differ from those of the others. Here, our formal modeling leads to a precise characterisation of the dynamic properties of pluralistic ignorance often reported by social scientists: its stability (if nothing special happens, the situation remains the same) and its extreme fragility (changing the behavior of one single agent might entirely turn around the situation).

Overall, by providing these two case studies of social phenomena, Part II shows how significant social phenomena can be specified in logical languages, how their information flow can be represented in terms of models for these languages, and how logic also helps to understand the theoretical core features of social scenarios that determine their limit behavior in the long run.

With our basic techniques for modeling stepwise updates in place, Part III turns to two further questions about network evolution. First, what are the logical properties of pure diffusion dynamics in social networks? And second, how does agents' knowledge of the network structure and of social influence effects interact with diffusion processes?

As for pure diffusion, we present two different takes: using a general dynamic framework based on hybrid logic in Chapter 5, and a simplified propositional dynamic logic in Chapter 7.

The general hybrid framework designed in Chapter 5 has the advantage of allowing any locally definable rule of influence to be "plugged-into" the logic, allowing for reasoning about a wide variety of diffusion phenomena. We show how to apply the logic to real-life documented phenomena: the previously studied case of pluralitic ignorance, and the diffusion of microfinance in villages.

To address the second question, Chapter 6 defines an epistemic extension of the preceding hybrid approach, and shows how agents might guess each other's private opinions by observing how their public behavior evolves in response to social conformity pressure. The resulting framework allows us to reflect how diffusion dynamics induce specific correlated learning dynamics.

Finally, using a further combination of our earlier logics, Chapter 7 shows how knowing more about the network structure and the behavior of agents in the network may accelerate diffusion in threshold models. We show how "smart" agents using all information available might anticipate diffusion. In the limit, when the network structure and agents behavior are common knowledge, the acceleration is maximal: the diffusion jumps directly to its fixed point in one step.

Finally, Part IV summarizes what we take to be the main conclusions from this thesis, presents some ongoing work, and points at some perspectives for future research that builds on the groundwork in this thesis.

1.4 Sources of the Chapters

Chapter 3 is based on:

A. Baltag, Z. Christoff, J.U. Hansen, & S. Smets, (2013). Logical Models of Informational Cascades. In J. van Benthem & F. Liu (Eds.) *Logic Across the University: Foundations and Applications*, vol. 47 of *Studies in Logic*, pp. 405–432. College Publications.

Chapter 4 is based on:

Z. Christoff & J.U. Hansen, (2013). A Two-Tiered Formalization of Social Influence. In D. Grossi, O. Roy, & H. Huang, (Eds.) *Logic, Rationality and Interaction, Proceedings of the 4th International Workshop (LORI 2013)*, vol. 8196 of *LNCS*, pp. 68–81. Springer.

Chapter 5 is based on:

Z. Christoff & J.U. Hansen, (2015). A Logic for Diffusion in Social Networks. *Journal of Applied Logic*, *13*(1), pp. 48–77. Elsevier.

Chapter 6 is based on:

Z. Christoff, J.U. Hansen, & C. Proietti (2016). Reflecting on Social Influence in Networks. To appear in *Journal of Logic, Language and Information*. Springer.

Chapter 7 is based on :

Z. Christoff, & R.K. Rendsvig (2014). Dynamic Epistemic Logic for Threshold Models. Presented at ELISIEM Workshop 2014, Tübingen.

A. Baltag, Z. Christoff, R.K. Rendsvig, & S. Smets (2015). Dynamic Epistemic Logic of Prediction and Diffusion in Threshold Models. ILLC Prepublication Series PP-2015-22. Submitted.

Chapter 2

Background Knowledge

In this chapter, we list some essential formal tools that will be used throughout the thesis, first from social network analysis, and then from logic. In addition to providing basic information, this will also serve to fix terminology for the main body of this dissertation. Readers who are familiar with all or part of this material can skip straightaway to the following Chapters.

2.1 Social Network Analysis

This section presents a few essential notions from social network analysis, including some simplifying assumptions in the literature that will be used throughout the thesis. Our presentation will be fast, since the different chapters to come will give more elaborate definitions where needed. The main source for the material presented here is [71].

As social networks can be represented by graphs, we start with a few graph theoretical notions that will recur in this thesis.

2.1.1 Graph Theory for Social Networks

First things first, a *graph* is a set of nodes, among which some pairs are connected by edges, and what we will call a *social network* is a graph where nodes are agents and edges represent some social relationship.

The following notions will come useful when describing local properties of a graph, i.e. the way two given nodes are related: A *path* in a graph is a sequence of nodes such that any two consecutive nodes are related by an edge. A *cycle* is a path such that the first and last node in the sequence are the same (and the sequence contains at least one different node). The *distance* between two nodes is the length of the shortest path between them, and two nodes are said to be n-*distant* if this path is of length $n + 1$.

On the global level, a graph is then said to be *connected* if there exists a path in between any two nodes, *n-connected* if the distance between any two nodes is at most n, and *fully connected* if any two nodes are related by an edge. The *diameter* of a graph is the smallest n such that the graph is n-connected. Moreover, a graph is said to be *acyclic* if it contains no cycle.

Throughout this thesis, we will impose the following constraints on social networks: they are *undirected, without self-loop, connected, and finite.* This corresponds to restricting ourselves to symmetric and irreflexive social relationships, such as friendship or neighborship, and to populations such that there exists a finite path in between any two pairs of agents, in other words, a "community". When two agents are related to each other by an edge in a social network, we will therefore often say that they are "neighbors", or "friends".[1]

The well-known *small-world phenomenon* refers to the fact that the diameter of big social networks is usually surprisingly short.[2]

However, to capture how connected a network is, it is not sufficient to describe its diameter. Indeed, for diffusion phenomena for instance, it is also important to be able to talk about how *dense* the network is. We will use the following notions:

The *clustering coefficient* of an agent is the proportion of her neighbors who are also neighbors to each other, and the clustering coefficient of a network is the average of the clustering coefficient of the agents in the networks. A *cluster of density d* is a set of agents such that, for each agent in the set, the proportion of her neighbors which are also in the group is at least d.[3] Note that an entire network (of at least two agents) is a cluster of density 1, and that, when assuming irreflexivity, a single agent is a cluster of density 0.

In the next section, we will briefly illustrate how the static network properties introduced above constrain the diffusion phenomena over networks.

2.1.2 Diffusion Phenomena

Throughout this thesis, we will assume that social networks do not change. What we are interested in capturing is another type of change: the way information, opinions, behavior, or trends, can spread over a given social network.

A typical example is the diffusion of a contagious disease within a population. Assuming that the disease can only be contracted by contagion from a neighbor in one's social network, the network structure (and in particular, the properties

[1] In Chapter 7, we will also require that networks are serial, i.e that each agent is related to an agent, a natural assumption when modeling diffusion processes, as an isolated unique agent would be an irrelevant case.

[2] The small-world phenomenon is sometimes confused with the famous but wrong idea that any real social network presents at most "six degrees of separation". See [p.30][71] for more details.

[3] Note that what is called a cluster of density d in [71] is called a *d-cohesive set* in [126].

introduced above) constrains how the infection will spread: how fast it spreads, and what would be needed to contain it.

Epidemic behavior is of course not restricted to diseases, it can be generalized to any feature of agents which may be affected by the features of their neighbors. For instance, the way in which we commonly describe how information circulates is directly inspired from this idea of the spreading of diseases (e.g. think of the spread of computer viruses), and social network analysis also studies for instance how some shared information "gets viral" within a community.

The "SI Model". Let us start by considering a minimal example: the diffusion of a disease within a population. Assume that each agent of the population is in either of two states: *infected* with the disease or *susceptible* to it. This type of models is commonly called "the SI Model" in the social networks literature [127]. Moreover, assume for instance that the disease can only be contracted by having a neighbor infected in one's social network graph.

Consider now how such an infection spreads within a community. This depends on how contagious the disease is. Assume for instance that each agent linked to an infected agent in the network will get infected too on the next day. This means that if some agent a is infected to start with, all agents directly linked to him will be infected on the next day, and then all agents linked to the agents linked to him, and so on. According to this rule of contagion (and our connectedness assumption), all agents will get sick at some point. The social network structure constrains how fast such a disease spreads. Locally, the *distance* from an agent b to the initially infected agent a determines how many days it takes before agent a will get infected. And globally, if nothing is done to stop the epidemics, the whole population will be infected after a number of days which is at most equal to the *diameter* of the network.

In this example, the dynamics is essentially captured by the following local diffusion rule: If any of your neighbors is infected, become infected yourself at the next moment (and stay infected forever).

However, one could very well imagine a different diffusion rule: after being infected, an agent immediately recovers and becomes susceptible again at the next moment. According to this new dynamic rule, agents might keep alternating forever between being infected and being susceptible and the network might never reach a stable state. Yet another type of dynamics is given by the "SIR Model" (susceptible, infected, recovered), where agents recover after being infected and are from then on immune to the disease.

In this thesis, we will focus on logics for diffusion processes of both types described above: where agents can "get infected" by the spreading feature and stay infected forever, in Chapter 7, and the case where they can "get infected" and get "disinfected", in Chapters 4 to 6.

Moreover, while the work done on simulations of networks is typically stochastic, we will restrict ourselves to diffusion rules which, as in the examples above, are deterministic. We believe that the limit case of deterministic diffusion rules is a good starting point for a logical approach. In every case modeled in this thesis, probabilities could be added, and in various ways.

In the next section, we will introduce a specific type of diffusion rules, where contagion occurs only when a certain *proportion* of neighbors are infected.

Threshold models. The notion of "threshold-limited influence" [71, 152], relies on a conformity pressure effect: *agents adopt a behavior, product, opinion, fashion, etc., whenever a critical fraction of their neighbors in the network have adopted it already.*

The so-called *threshold models*, first introduced by [90, 144], are used precisely to represent the dynamics of diffusion under threshold-limited influence. This type of models has received a lot of attention in recent literature [71, 109, 126, 148, 8, 87, 118, 121]. Chapter 7 will design a logical framework for the diffusion dynamics in threshold models.

The simplest type of threshold model consists of three components: a fixed social network, a set of agents in the network which are "infected" (hold an opinion, follow a fashion, etc), and a fixed uniform adoption threshold indicating what proportion of neighbors need to be infected for the infection to spread. A threshold model thus represents the current spread of an infection over the network, while containing the adoption threshold which prescribes how this spread will evolve. In such a model, infected agents will stay infected forever.

We have illustrated above how a static property of the network, its diameter, determines an upper bound on the time left before an entire population would get infected in the SI model.[4] Similarly, we can illustrate how the density of a network constrains the diffusion in threshold models. The higher the density of a group, the better it will resist to influence from the outside [71, Ch.19.3]. The following result from [126] makes this precise: for a given threshold θ, assuming that some agents are already infected, the infection will reach the entire population if and only if there is no cluster of density $1-\theta$ among the healthy agents.

In the above, we have introduced a few notions from graph theory and social network analysis which we will rely on throughout this thesis. We have given examples of models of diffusion phenomena and illustrated how the static graph-theoretical notions introduced constrain the diffusion dynamics we are interested in. We will encounter more examples of this sort on our way, in particular in Chapter 4, 7, and 8.

In the next section, we will introduce the logical background which we take as our starting point in our quest for a logic for information and diffusion dynamics

[4]Note that the SI model above is a specific case of threshold model, where the threshold is set to be $\frac{1}{n}$ and n is at least as big as the maximal number of neighbors of any agent in the network.

on social networks.

2.2 Logic

This section introduces the logical background from this thesis. We first introduce tools from dynamic epistemic logic [24, 70, 47], used to model knowledge and beliefs of agents and their dynamics upon receiving new information. Later on, we give a brief presentation of the existing work by Seligman, Liu and Girard [145, 166, 120], on logics combining social networks to knowledge and belief.

2.2.1 Dynamic Epistemic Logic

The work on dynamic epistemic logic (DEL) [24, 70, 47] brings together two structural ingredients, i.e. epistemics and dynamics, in one unified setting. On the one hand, DEL allows us to model the static attitudes of different agents, be it knowledge, beliefs, preferences or another attitude. On the other hand, it allows us to model the change of these attitudes, i.e. knowledge updates, belief dynamics and change of preferences.

In this section, we present the basic notions from the logic of knowledge update and belief change. To do so, we will use the definitions from the work on epistemic logic, doxastic logic and dynamic epistemic logic [47, 41, 70, 24, 27, 123, 76]. Moreover, our presentation will follow very closely the order and style of presentation adopted in [81, Section 2.2].

Epistemic Logic. We will start here by introducing the very basic components of epistemic logic [97]. The epistemic language contains a modality K_a for each agent $a \in \mathcal{A}$.

2.2.1. Definition. [Syntax of $\mathcal{L}_{\mathrm{K}}$] The epistemic language $\mathcal{L}_{\mathrm{K}}$ is defined as follows:

$$\varphi := \ p \mid \neg\varphi \mid \varphi \vee \varphi \mid K_a\varphi$$

where $p \in \Phi$, $a \in \mathcal{A}$, $\mathcal{A}$ is a finite set of agents, and Φ is a countable set of atomic sentences.

Besides the standard Boolean operators, this language contains the epistemic construction $K_a\varphi$ which we read as "agent a knows that φ".

To build an interpretation, we first introduce the concept of an epistemic (state) model, given by a set of possible worlds and, for each agent a in a given finite set $\mathcal{A}$, a binary relation, representing agent a's subjective epistemic indistinguishability:

2.2.2. Definition. [epistemic (state) model] An epistemic (state) model $\mathcal{M}$ is a triple:

$$(W, (\sim_a)_{a\in\mathcal{A}}, V),$$

where $W \neq \emptyset$ is a set of possible worlds, for each $a \in \mathcal{A}$, $\sim_a$ is a binary equivalence relation on W, and $V : \Phi \to \mathcal{P}(W)$ is a valuation.

An agent's subjective epistemic indistinguishability is here represented via an equivalence relation. In an alternative representation, following e.g. [13], this relation defines separate information cells: the information cell $[w]_a$ of agent a in state w is the set of worlds $v \in W$ such that $w \sim_a v$. What agent a knows in world w is therefore defined as what is true everywhere in $[w]_a$:

2.2.3. Definition. [Semantics of $\mathcal{L}_{\mathrm{K}}$] Let $\mathcal{M} = (W, (\sim_a)_{a\in\mathcal{A}}, V)$, $w \in W$, $p \in \Phi$ and $\varphi \in \mathcal{L}_{\mathrm{E}}$. The truth of φ at world w in $\mathcal{M}$ is defined as follows:

$$\begin{array}{lll}
\mathcal{M}, w \models p & \text{iff} & w \in V(p) \\
\mathcal{M}, w \models \neg\varphi & \text{iff} & \text{it is not the case that } \mathcal{M}, w \models \varphi \\
\mathcal{M}, w \models \varphi \vee \varphi & \text{iff} & \mathcal{M}, w \models \varphi \text{ or } \mathcal{M}, w \models \psi \\
\mathcal{M}, w \models K_a\varphi & \text{iff} & \text{for all } v \text{ such that } w \sim_a v \text{ we have } \mathcal{M}, v \models \varphi
\end{array}$$

The proof system of epistemic logic K is axiomatized by using the axioms of S5 and the rule of modus ponens and necessitation below:

$$\begin{array}{ll}
\text{PL} & \vdash \varphi \text{ if } \varphi \text{ all instantiations of propositional tautologies} \\
\text{K} & \vdash K_a(\varphi \to \psi) \to (K_a\varphi \to K_a\psi) \\
\text{T} & \vdash K_a\varphi \to \varphi \\
4 & \vdash K_a\varphi \to K_aK_a\varphi \\
5 & \vdash \neg K_a\varphi \to K_a\neg K_a\varphi \\
\text{MP} & \text{if } \vdash \varphi \to \psi \text{ and } \vdash \varphi, \text{ then } \vdash \psi \\
\text{Nec} & \text{if } \vdash \varphi, \text{ then } \vdash K_a\varphi
\end{array}$$

Now that we have defined a static multi-agent epistemic model, let us turn to how the knowledge state of an agent changes upon receiving new information. We start with the simplest (and most radical) case of model transformation: update.

Update. The model transformation operation which incorporates the new information that φ is the case, when received from a source which is considered to be infallible, is called "update". Updating with formula φ simply deletes all possible worlds which did not satisfy φ.

2.2.4. Definition. [update]

The update of an epistemic model $\mathcal{M} = (W, (\sim_a)_{a\in\mathcal{A}}, V)$ with a formula φ, restricts $\mathcal{M}$ to those worlds that satisfy φ, formally $\mathcal{M} \mid \varphi = \mathcal{M}' := (W', (\sim'_a)_{a\in\mathcal{A}}, V')$ is given by:

$$\begin{array}{l}
W' = \{w \in W \mid w \models \varphi\}; \\
\text{for each } a \in \mathcal{A}, \sim'_a = \sim_a\restriction W'; \text{ and} \\
V' = V\restriction W'.
\end{array}$$

The event of announcing truthful information to all agents is called a "public announcement" and is noted by $!\varphi$. The language of public announcement logic (PAL) $\mathcal{L}_{PAL}$ consists of the language $\mathcal{L}_K$ above, extended with dynamic formula $[!\varphi]\psi$, reading "after public announcement of φ, ψ". One of the general strengths of DEL is that it can talk about the communication events themselves. Since φ can itself contain dynamic modalities, the language of PAL can talk about public announcements about public announcements, for instance.

2.2.5. Definition. [Semantics of $\mathcal{L}_{PAL}$] The semantics of $\mathcal{L}_{PAL}$ is obtained by the semantics of $\mathcal{L}_E$ extended with the following clause:

$$\mathcal{M}, w \models [!\varphi]\psi \quad \text{iff} \quad \text{if } \mathcal{M}, w \models \varphi \text{ then } \mathcal{M} \mid \varphi, w \models \psi$$

The proof system of the logic PAL is obtained by adding the following "reduction axioms" [133] to the epistemic logic K given above:

$$\begin{array}{l} \vdash [!\varphi]p \leftrightarrow (\varphi \to p), \text{ for } p \in \Phi \\ \vdash [!\varphi]\neg\psi \leftrightarrow (\varphi \to \neg[!\varphi]\psi) \\ \vdash [!\varphi](\psi \vee \xi) \leftrightarrow ([!\varphi]\psi \vee [!\varphi]\xi) \\ \vdash [!\varphi]K_a\psi \leftrightarrow (\varphi \to K_a[!\varphi]\psi) \end{array}$$

The logic PAL is complete with respect to the class of epistemic models [133]. Beyond public announcements, let us now generalize to other epistemic events.

Product update. An essential part of DEL is the mechanism that it uses to represent event models and to let an event model act on a given state model: this mechanism is standardly called the "product update" [24]. The product update of an event model and a given state model produces a new state model. In effect it means that the dynamic dimension in DEL expresses a model transforming operation, an operation which marks a crucial difference with the representation of actions (as a relation over possible worlds) in labelled state transition systems such as e.g., propositional dynamic logic (PDL) [95].

2.2.6. Definition. [multi-agent event model] A multi-agent event model is a triple:

$$\mathcal{E} = (E, (\sim_a^{\mathcal{E}})_{a\in\mathcal{A}}, \mathtt{pre}),$$

where $E \neq \emptyset$ is a set of events, $\sim_a^{\mathcal{E}}$ is a binary equivalence relation on E, and $\mathtt{pre} : E \to \mathcal{L}_{EL}$, is a precondition function where $\mathcal{L}_{EL}$ is a given epistemic language.

The relation $\sim_a^{\mathcal{E}}$ encodes what the agent a knows about the event itself. An event can only occur in a state if that state satisfies its precondition.

In order to model fact changing operations in DEL via "actions which change the truth values of atomic sentences" [24], one has to adapt the logical setting as in e.g. [24, 47, 17]. One standard way of doing this, is by extending the event models with the specification of a postcondition to indicate the new valuation after the event has happened. We will use DEL with fact changing operations later in this thesis.

Given an initial state model and an event model, the product update yields a general way for computing the epistemic state resulting from the given event applied to the given state model [24]. The landscape of DEL comprises a whole range of logical systems that all use the same idea of a model transforming mechanism (although the actual update mechanism itself can vary). The above public announcement logic is seen as a special case of the following more general mechanism which allows for both public and private actions between agents:

2.2.7. Definition. [product update] Let $\mathcal{M} = (W, (\sim_a)_{a\in\mathcal{A}}, V)$ be an epistemic model and $\mathcal{E} = (E, (\sim_a^{\mathcal{E}})_{a\in\mathcal{A}}, \mathtt{pre})$ an event model. The *product update* $\mathcal{M} \otimes \mathcal{E}$ states of which are the pairs (w, e) such that $w \in W$, $e \in E$, and w satisfies the precondition of e. The epistemic relation in the resulting model is defined as:

$$(w, e) \sim_a (w', e') \text{ iff } w \sim_a w' \text{ and } e \sim_a^{\mathcal{E}} e',$$

and the valuation is as follows:

$$(w, e) \in V(p) \text{ iff } w \in V(p).$$

Belief. We now turn to an extended language to capture both knowledge and belief. The following doxastic-epistemic language is obtained by adding a belief modality to language $\mathcal{L}_K$:

2.2.8. Definition. [Syntax of $\mathcal{L}_{KB}$] The syntax of doxastic-epistemic language $\mathcal{L}_{\mathrm{KB}}$ is defined as follows:

$$\varphi := \ p \mid \neg\varphi \mid \varphi \vee \varphi \mid K_a\varphi \mid B_a^{\psi}\varphi$$

where $p \in \Phi$ is an element of a given countable set atomic sentences and $a \in \mathcal{A}$ is an element of a given finite set of agents.

In addition to the standard Boolean operators, we have included a conditional belief operator $B_a^{\psi}\varphi$, which intuitively reads as "conditional on receiving ψ, agent a would believe that φ was the case". The temporal element in this interpretation is crucial, as explained in [27]. Note further that belief simpliciter can be defined in terms of the conditional belief operator by setting $B_a\varphi := B_a^{\top}\varphi$ as given by conditioning on a tautology.

To give an interpretation to the language construction containing the belief modality, we introduce a new type of models. Different from the standard KD45 models in modal logic, often used to represent the doxastic attitudes of agents, we work with so-called epistemic-plausibility models, following the work in [28, 27, 42]. The following definitions are given in [27]:

2.2.9. Definition. [epistemic-plausibility model] An *epistemic-plausibility model* $\mathcal{M}$ is a triple

$$(W, (\sim_a)_{a\in A}, (\leq_a)_{a\in A}, V),$$

where $W \neq \emptyset$ is a set of states, for each $i \in A$, $\leq_i$ is a total well-founded preorder on W, and $V : \Phi \to \mathcal{P}(W)$ is a valuation.

A pair $(\mathcal{M}, w)$, where $\mathcal{M} = (W, (\sim_a)_{a\in A}, (\leq_a)_{a\in A}, V)$ an epistemic plausibility model and $w \in W$, is called a *pointed epistemic plausibility model.*

For each $a \in A$ we will assume that $\leq_a \subseteq \sim_a$.[5]

2.2.10. Definition. [Semantics of $\mathcal{L}_{KB}$] The truth of a formula $\varphi \in \mathcal{L}_{KB}$ in a model $\mathcal{M} = (W, (\sim_a)_{a\in A}, (\leq_a)_{a\in A}, V)$ is given by the epistemic clauses given above extended with the following clause:

$$\mathcal{M}, w \models B_a^{\psi}\varphi \quad \text{iff} \quad \text{for all } v \in [w]_a \ \text{ if } \ v \in \min_{\leq_i}([w]_a \cap \|\psi\|) \text{ then } v \models \varphi$$

The proof system of $\mathcal{L}_{KB}$ and its complete axiomatization is given in [27].

Plausibility upgrade. Now that we have seen how to model beliefs in terms of a plausibility ordering, let us introduce ways to represent belief change[6] when the agents face new incoming information. If the information source is taken to be less than infallible, a "softer" type of model transformations can be applied, which will result in a reordering of the possible states in the model, but will not delete any [27, 42]. There is no unique way of defining such a reordering.

For simplicity, we will only introduce one example of such a plausibility upgrade, the "lexicographic upgrade".[7]

The effect of a lexicographic upgrade with formula φ is to force all worlds which satisfied φ before the announcement to become more plausible than the ones which did not. We use the following notation for the plausibility order relative to satisfaction of a formula: $\leq_a^{\varphi} = \leq_a\restriction \|\varphi\|$, and $\leq_a^{\bar{\varphi}} = \leq_a\restriction \|\neg\varphi\|$.

2.2.11. Definition. [lexicographic upgrade] The *lexicographic upgrade* of an epistemic plausibility model $\mathcal{M} = (W, (\sim_a)_{a\in A}, (\leq_a)_{a\in A}, V)$ with a formula φ is defined as follows:

$$\mathcal{M} \Uparrow \varphi := (W, (\sim_a)_{a\in A}, (\leq'_a)_{a\in A}, V),$$

[5] Note that with this assumption, the definition of a plausibility model can be simplified to $(W, (\leq_a)_{a\in A}, V)$, when defining the epistemic accessibility relation in terms of the given plausibility relation, as is done in [27].

[6] Note that our notion of "belief change" refers to the work in dynamic epistemic logic where one makes a distinction between static AGM belief revision (modelled via conditional modal operators) and dynamic belief change (modelled via dynamic modal operators). This distinction is in detailed explained in [27].

[7] We leave out the notions of "radical upgrade" and "conservative upgrade" in [27]. Note that under various names, such upgrades have been previously proposed in the literature on Belief Revision, e.g. by Rott [137], and in the literature on dynamic semantics for natural language by e.g. Veltman [159].

where for each $a \in A$ and for all $v, w \in [w]_a$:

$$v \leq'_a w \text{ iff } (v \leq^{\varphi}_a w \text{ or } v \leq^{\bar{\varphi}}_a w \text{ or } (v \models \varphi \text{ and } w \models \neg\varphi)).$$

The language of announcements inducing lexicographic upgrade is defined in the following way.

2.2.12. Definition. [Syntax of $\mathcal{L}_{KB\Uparrow}$] The syntax of the doxastic-epistemic language $\mathcal{L}_{KB\Uparrow}$ is defined by extending $\mathcal{KB}$ with $[\Uparrow\varphi]\psi$.

The semantics is obtained by adding the following clause to the previously defined semantics:

2.2.13. Definition. [Semantics of $\mathcal{L}_{KB\Uparrow}$] Let $\mathcal{M}$ be a doxastic model, $w \in W$ the truth of a formula $\phi \in \mathcal{L}_{KB\Uparrow}$ is given by the semantic clauses for $\mathcal{L}_{KB}$ extended with:

$$\mathcal{M}, w \models [\Uparrow\varphi]\psi \quad \text{iff} \quad \mathcal{M} \Uparrow \varphi, w \models \psi$$

The given lexicographic upgrade mechanism is limited in the sense that it assumes that all agents in the model will upgrade their beliefs in the same way. Or in other words, it is common knowledge that all agents consider the source fallible but highly trusted in the same way. In many scenario's this assumption is unrealistic and hence it has to be dropped. Different agents can have different levels of trust in the information they receive. In addition some information may be privately available only to a subgroup of agents. To model these more advanced scenarios, we again have to turn to the standard ingredients of DEL by including event models and a general upgrade mechanism. In the context of belief change, the product update mechanism has to be adapted to allow our agents to handle "soft information". The mechanism in [27] designed for this task is called the "action-priority update rule". For the details of this construction, we refer the reader to [27, 44].

2.2.2 Facebook Logic

Aside from DEL, we rely on another line of work within the field of logic, the two-dimensional hybrid logic developed by Seligman, Girard, and Liu. Their three initial papers in this direction propose such a framework to capture three types of changes in social networks: the initial "Facebook logic" framework [145], dealing with *knowledge* change via communication through an (online) social network; and two variations of the first one, one of which applies to *preferences* change under peer pressure [166], the other one to *belief* change under social influence [120]. We present below the main ingredients of the latter.

The novelty of this line of work is to allow the logic to talk about both the attitudes of agents (knowledge, preferences, or opinions) and about the social network structures. Formally, this is rendered by a two-dimensional setting, where one dimension represents the (knowledge, preferences, or opinion) state of the agents, and the other dimension captures the social network structure.

Below, we introduce the formal tools from the epistemic setting of [145], followed by the ideas on opinion change under peer pressure from [120]

Epistemic social network logic. The "Facebook logic" from [145] is obtained by combining a social network dimension to an epistemic dimension. More precisely, an epistemic network model is a multi-agent epistemic model, as introduced in Section 2.2.1, with a social network structure in each possible world:

2.2.14. Definition. [Epistemic Social Network Model]

A *model* is a tuple

$$\mathcal{M} = (A, W, (\asymp_w)_{w \in W}, (\sim_a)_{a \in A}, g, V),$$

where: A is a non-empty set of agents, W is a non-empty set of possible worlds, $\asymp_w$ is an irreflexive, symmetric, binary relation on A, for each $w \in W$ (representing the network structure at the world w), $\sim_a$ is an equivalence relation on W for each $a \in A$ (representing the uncertainty of a), $g : \mathsf{NOM} \to A$ is a function assigning an agent to each nominal, and $V : \Phi \to \mathcal{P}(W \times A)$ is a valuation.

In addition to the standard knowledge modality K, the language includes tools from hybrid logic [10, 57]: *Nominals* are used to refer to the agents, and operators $@_i$, to switch the evaluation point to the unique agent named by nominal i. The semantics is therefore indexical, in the sense that a formula is evaluated both at a world and at some given agent. An additional modal operator F quantifies over friends (or network neighbors): F reads "all of my friends" and its dual, $\langle F \rangle$, "some of my friends".

2.2.15. Definition. [Epistemic Social Network Syntax]

The syntax of the epistemic network language $\mathcal{L}$ is defined as follows[8]:

$$\varphi \quad ::= \quad p \mid i \mid \neg\varphi \mid \varphi \wedge \varphi \mid F\varphi \mid @_i\varphi \mid K\varphi \mid \downarrow i\varphi \mid$$

where $p \in \Phi$, and $i \in \mathsf{NOM}$.

The semantics is defined as follows:

[8][145] discusses some extensions of this language with indexical announcement, which we do not include in this introduction.

2.2.16. Definition. [Epistemic Social Network Semantics]

Given a model $\mathcal{M} = (A, W, (\asymp_w)_{w\in W}, (\sim_a)_{a\in A}, g, V)$, $a \in A$, $w \in W$, and formulas $p \in \Phi, i \in \mathsf{NOM}$, *and* $\varphi \in \mathcal{L}$, the truth of φ at (w, a) in $\mathcal{M}$ is given by:

$$
\begin{array}{lll}
\mathcal{M}, w, a \models p & \text{iff} & (w, a) \in V(p) \\
\mathcal{M}, w, a \models i & \text{iff} & g(i) = a \\
\mathcal{M}, w, a \models \neg\varphi & \text{iff} & \text{it is not the case that } \mathcal{M}, w, a \models \varphi \\
\mathcal{M}, w, a \models \varphi \wedge \psi & \text{iff} & \mathcal{M}, w, a \models \varphi \text{ and } \mathcal{M}, w, a \models \psi \\
\mathcal{M}, w, a \models F\varphi & \text{iff} & \forall b \in A; a \asymp_w b \Rightarrow \mathcal{M}, w, b \models \varphi \\
\mathcal{M}, w, a \models K\varphi & \text{iff} & \forall v \in W; w \sim_a v \Rightarrow \mathcal{M}, v, a \models \varphi \\
\mathcal{M}, w, a \models @_i\varphi & \text{iff} & \mathcal{M}, w, g(i) \models \varphi \\
\mathcal{M}, w, a \models \downarrow i\varphi & \text{iff} & \mathcal{M}, w, g^i_a \models \varphi \\
 & & \text{where } g^i_a = y \text{ if } i = y \text{ and } g(y) \text{ otherwise.}
\end{array}
$$

We will use the same hybrid tools to describe the network structure in Chapters 4 to 6. Moreover, we will use an extension of this two-dimensional epistemic hybrid setting in Chapter 6 to model the interaction between knowledge and diffusion phenomena.

In addition to the formal tools from [145] introduced above, we also build on the main ideas of the work of Seligman, Girard, and Liu on diffusion of opinions under social influence from [120]. We will present those ideas below.

Opinion Change under Peer Pressure. The setting of [120] assumes that each agent is always in one of the three following doxastic states, relatively to a given proposition φ: either she believes that φ ($B\varphi$), or she believes that $\neg\varphi$ ($B\neg\varphi$), or she is undecided about φ: ($U\varphi$ – an abbreviation of $\neg B\varphi \wedge \neg B\neg\varphi$). Sentences are interpreted indexically at an agent in this hybrid setting: if p means "I am a logician", BFp reads "I believe that all my friends are logicians" and FBp reads "each of my friends believes that s/he is a logician".

This static framework is combined with an influence operator to represent how belief repartition changes in a community, according to the following peer pressure principle: every agent tends to align her belief with the ones of her friends. The notions of Strong Influence and Weak Influence are defined, corresponding respectively to the belief changing operators of revision (adoption of an opinion) and contraction (abandon of an opinion) in the tradition of [4]. These two notions determine entirely the dynamics:

An agent is *strongly influenced* (SI) to believe φ when *all* of her friends (and at least one) believe that φ:

$$SI\varphi := FB\varphi \wedge \langle F\rangle B\varphi$$

An agent under strong influence with respect to φ will come to believe φ too whatever her initial attitude towards φ was[9]. An agent is already *weakly influ-*

[9]This implies that the revision process is restricted to successful formulas

enced (WI) with respect to φ when *some* of her friends believe that φ and none of her friends believe that $\neg\varphi$:

$$WI\varphi := F\neg B\neg\varphi \wedge \langle F\rangle B\varphi$$

Under weak influence, if the agent was undecided or if she already believed that φ, nothing changes; but if she believed that $\neg\varphi$, she will drop her belief and become undecided.

This simple setting is sufficient to model how opinions (about an implicit φ) will spread in a community. Moreover, it allows for a simple characterization of the stability and stabilization conditions of the influence process.

Stabilization. When is a configuration of opinions stable, according to the above-defined dynamic rule? It is stable, trivially, when all agents are already in the same state, but this is not the only case. For instance, if each agent has both some friends who believe φ and some who believe $\neg\varphi$, or only has undecided friends, then nothing will happen.

As in the case of the above mentioned threshold models, let us remark that some configurations are particularly resistant to change: for example, a unanimously undecided (and therefore stable) community of three friends in which one agent would start believing φ (for any reason other than influence) would continue to be stable; and if all agents initially believe that φ and one agent revises with $\neg\varphi$, the community will resist the change even more strongly and immediately go back to its initial (stable) state.

Interestingly, some configurations will never stabilize. If, for each agent (who has at least one friend) in the community, all of her friends believe that φ and all of their friends believe that $\neg\varphi$, they will keep switching beliefs forever, stuck in a loop. Some configurations will stabilize after some iterated influence changes: one undecided agent is sufficient for a configuration to be becoming stable.

One elegant advantage of such a simple framework is that it can talk about stability and about stabilization within the language of friendship and belief. The sufficient and necessary conditions simply correspond to the negation of the preconditions of the revision and contraction actions, namely, SI and WI, both for a given formula and for its negation.[10] Moreover, all configurations which are *becoming stable* can be fully characterized too, by simply negating the description "looping scenario" mentioned above.[11][12]

[10]For details, see [120][p. 8].

[11]Note that the only oscillating configurations are the ones where $B\varphi$ and $B\neg\varphi$ are distributed as a proper 2-coloring of the network graph. As a consequence, all networks which are not 2-colorable, i.e. all networks containing cycles of odd length, guarantee stabilization of the opinion distribution. We will come back to 2-colorability in Chapters 4 and 8.

[12]For more details about stabilization conditions for different diffusion rules and expressivity of the corresponding languages, see Chapter 8 and our ongoing work in [61].

This thesis takes its inspiration directly from the preceding lines of work, as it investigates the way opinions and knowledge will or should change under social influence.

More concretely, the material introduced in this chapter allows us to (1) describe properties of social network structures using graph theoretical notions, (2) describe diffusion phenomena using models from social network analysis, (3) model beliefs and knowledge of agents, and how these change when new information is received using techniques from dynamic epistemic logic, and (4) specify and axiomatize reasoning about social network structure and the evolution of opinion under peer pressure, using ideas from Facebook logic and its follow-up literature.

In the coming chapters, we will combine and enrich all those tools in several ways as we confront concrete social scenarios and identify major general phenomena to be explained in their proper generality.

Part II

Modeling Collective Failures

Introduction to Part II: From Micro Success to Macro Failure?

Individual preferences sometimes give rise to counterintuitive results at the collective level. This discrepancy between *microbehavior* and *macrobehavior*, illustrated by Schelling's work [144, 142, 143], is well-established in social sciences and economics:

> Economists are familiar with systems that lead to aggregate results that the individual neither intends nor needs to be aware of, the results sometimes having no recognisable counterpart at the level of the individual.[142, p.488]

A celebrated example is Schelling's work on segregation modeling, thanks to which he shows that strongly segregated neighborhoods are surprisingly not the result of a strong preference of its habitants for living among similar individuals [142, 143].

This part of the thesis presents a logical case study of two social-epistemic phenomena where individual and group behaviors seem to mismatch in a similarly surprising way: individuals rationally hold beliefs which collectively lead the group to some level of epistemic failure.

Chapter 3, based on the work published in [18], models the herding dynamic phenomenon of *informational cascades*, where individuals are sequentially lead to rationally form the same belief, whether it is true or not.

Chapter 4, presenting work published in [62], focuses on another form of collective error: *pluralistic ignorance*, a situation where all agents hold the same belief but are all similarly mistaken about each other's beliefs, this uniformity leading everybody to stay in their state of error. Both phenomena feature discrepancies between individual rationality and collective epistemic felicity.

Modeling these two phenomena should be seen a first step towards giving a formal answer to a much more general question: under which conditions does micro success lead to macro failure?

Chapter 3

Informational Cascades

3.1 Introduction

Informational cascades are social-informational herding phenomena, in which sequential inter-agent communication might lead to collective epistemic failure, despite availability of information that should be sufficient to track the truth. This chapter models an example of an informational cascade in order to check the correctness of the individual reasoning involved.

As there is no consensus on what this “correctness” amounts to exactly, we will model the occurrence of the same cascade under two different rationality assumptions : 1) standard Bayesian rationality, where agents are perfect probabilistic reasoners; and 2) a simpler sumerical notion, where being rational amounts to simply being able to count and compare the number of pieces of evidence for two alternatives. To make sure cascades do not arise as the result of a lack of higher-order reasoning, we assume, in both cases, that agents also are perfect higher-order reasoners.

3.1.1 Outline

Section 3.1.2 first briefly introduces the phenomenon of informational cascades, and Section 3.1.3 presents our motivation for a logical modeling of these. Section 3.2 then recalls a paradigmatic example of a cascade and its standard Bayesian analysis, as given in [71]. We then model the same example using two different logical settings: an adaptation of probabilistic dynamic epistemic logic [48, 112] in Section 3.3, and a new framework for counting evidence in Section 3.4. Unlike the Bayesian modeling, both logical modelings represent agents with unbounded higher-order reasoning powers. Moreover, the first approach allows us to still model perfect probabilistic reasoners, while the second one allows us to model agents who use a heuristic reasoning method: they simply count pieces of evidence supporting each alternative. Thanks to this twofold logical analysis, we will be able to generalize, beyond Bayesian rationality, the result that cascades are sometimes unescapable by "rational" means. Indeed, the three different models confirm that the group's inability to track the truth is sometimes a direct consequence of each agent's rational attempt at truth-tracking. In other words, collective failure sometimes results from individual rationality.

3.1.2 The Phenomenon of Informational Cascades

We say that an informational cascade occurs when individuals in a sequence imitate the choices of others despite the fact that their own private information suggests otherwise. Let us illustrate the phenomenon by the example of the choice of a restaurant, as given in [71]: Suppose an agent has some private information that restaurant A is better than restaurant B. Nevertheless, when arriving at the adjacent restaurants she sees a crowded restaurant B and an empty restaurant A, which makes her decide to opt for restaurant B. In this case our agent interprets the others' choice for B as conveying some information about which restaurant is better and this overrides her independent information that restaurant A is better. However, it could very well be that all the people in restaurant B were confronted with the same choice and chose that restaurant for the exact same reason.

By observing the decisions of the previous people in a sequence, individuals in a cascade form an opinion about the information that the others might have, and this opinion may outweigh other (private) information. In this way, individuals in a sequence might be led to ignore their own private evidence and to simply start following the crowd, whether the crowd is right or wrong. Examples of informational cascades include bestseller lists for books, judges voting publicly and sequentially, peer-reviewing processes, etc. [51]. While models of such phenomena were independently developed in [50] and [34], the term *informational cascades* is due to [50]. Traditionally investigated by the social sciences, informational cascades have recently become subject of philosophical reflection, as part of the field of *Social Epistemology* [84, 86]. In particular, [99] gives a philosophical discussion of informational cascades and the more general class of "info-storms", their triggers and their defeaters ("info-bombs"), as well as the epistemological issues raised by the existence of such social-epistemic phenomena. We here take the first step towards an epistemic logical study of informational cascades by modeling cascades in a logical setting based on (both probabilistic and more qualitative) versions of Dynamic-Epistemic Logic.

3.1.3 Mindless Imitation?

Intuitively, it may seem like individuals in a cascade, when following the crowd in its error, must suffer from some form of irrationality: they must be misinterpreting how previous agents have formed their beliefs or they must lack reasoning power. However, it has already been shown that when agents are assumed to be

Bayesian reasoners, they may still all enter a cascade, despite the fact that each agent's opinion is perfectly justified, given the information that is available to her. In Section 3.2, we will recall one of the standard examples of such a "rational" cascade, the "Urn example", and its Bayesian model, as given in [71]. The inescapable conclusion seems to be that, in some cases, *individual rationality leads to group "irrationality"*.

However, what is typically absent from this standard Bayesian analysis of informational cascades is the agents' higher-order reasoning about other agents' minds and about the whole sequential protocol in which they are participating. So one may still argue that by such higher-order reflection (and in particular, by becoming aware of the dangers inherent in the sequential deliberation protocol), "truly rational" agents should be able to avoid the formation of cascades.

We will go one step further and prove that there exists situations in which no amount of higher-order reflection and meta-rationality can stop a cascade. To show this, we present (Section 3.3) a formalization of the above-mentioned Urn example using Probabilistic Dynamic Epistemic Logic [48, 112]. Epistemic logic takes into account all the levels of mutual belief/knowledge (beliefs about others' beliefs, etc) about the current *state* of the world; while dynamic epistemic logic adds also all the levels of mutual belief/knowledge about the on-going *informational events* ("the protocol"). As we will show, a cascade can still form. This proves our point: some cascades cannot be prevented even by the most perfect, idealized kind of individual rationality, one endowed with unlimited higher-level reflective powers. Informational cascades of this "super-rational" kind can be regarded as "epistemic Tragedies of the Commons": *paradoxes of (individual-versus-social) rationality.*

We then address a second objection raised against the Bayesian analysis of informational cascades (Section 3.4). One may argue that real agents, although engaging in cascades, do it for non-Bayesian reasons: instead of probabilistic conditioning, they seem to use "rough-and-ready" qualitative heuristic methods, e.g. by simply *counting* the pieces of evidence in favor of one hypothesis against its alternatives. To model the reasoning produced by this kind of qualitative reasoning (by agents who still maintain their higher-level awareness of the other agents' minds), we introduce a new framework – a *multi-agent logic for counting evidence.* We show that, even if agents only use such a less sophisticated heuristic way of reasoning than full-fledged probabilistic logic, they may still "rationally" engage in informational cascades. Hence, the above conclusion can now be extended to a wider range of agents: *as long as the agents can count the evidence, then no matter how high or how low are their reasoning abilities* (even if they are capable of full higher-level reflection about others' minds, or dually even if they can't go beyond simple evidence counting), *individual rationality may still lead to group "irrationality"*.

3.2 The Bayesian Analysis of Cascades

The section recalls the Bayesian analysis of an informational cascade. We will focus on a simple example that was created for studies of informational cascades in a laboratory [5, 6].

3.2.1 The Urn Example

Consider two urns, respectively named U_W and U_B, where urn U_W contains two white balls and one black ball, and urn U_B contains one white ball and two black balls. One urn is randomly picked (say, using a fair coin) and placed in a room. This setup is common knowledge to a group of agents, which we will denote $a_1, a_2, ..., a_n$ but they do not know which of the two urns is in the room. The agents enter the room one at a time; first a_1, then a_2, and so on. Each agent draws one ball from the urn, looks at it, puts it back, and leaves the room. Hence, only the person in the room knows which ball she drew. After leaving the room she makes a guess as to whether it is urn U_W or U_B that is placed in the room and writes her guess on a blackboard for all the other agents to see. Therefore, each individual a_i knows the guesses of the previous people $a_1, a_2, ..., a_{i-1}$ in the sequence $a_1, a_2, ..., a_n$ before entering the room herself. It is common knowledge that they will be *individually rewarded if and only if their own guess is correct.*

3.2.2 Bayesian Analysis of the Example

To present the Bayesian Analysis of the urn example, we will follow the presentation of [71]. Let us assume that in fact urn U_B has been placed in the room. When a_1 enters and draws a ball, there is a unique simple decision rule she should apply: if she draws a white ball it is rational to make a guess for U_W, whereas if she draws a black one she should guess U_B. We validate this by calculating the probabilities. Let w_1 denote the event that a_1 draws a white ball and b_1 denote the event that she draws a black one. The proposition that it is urn U_W which is in the room will be denoted similarly by U_W and likewise for U_B. Given that it is initially equally likely that each urn is placed in the room the probability of U_W is $\frac{1}{2}$ ($P(U_W) = \frac{1}{2}$), and similarly for U_B. Observe that $P(w_1) = P(b_1) = \frac{1}{2}$. Assume now that a_1 draws a *white* ball. Then, via Bayes' rule, the posterior probability of U_W is

$$P(U_W|w_1) = \frac{P(U_W) \cdot P(w_1|U_W)}{P(w_1)} = \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Hence, it is indeed rational for a_1 to guess U_W if she draws a white ball (and to guess U_B if she draws a black ball). Moreover, when leaving the room and making a guess for U_W (resp. U_B), all the other individuals can infer that she drew a white (resp. black) ball.

When a_2 enters the room after a_1, she knows the color which a_1 drew and it is obvious how she should guess if she draws a ball of the same color. If a_1 drew a white ball and a_2 draws a white ball, then a_2 should guess U_W. Formally, the probability of U_W given that both a_1 and a_2 draw white balls is

$$P(U_W|w_1, w_2) = \frac{P(U_W) \cdot P(w_1, w_2|U_W)}{P(w_1, w_2)} = \frac{\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{2}{3}}{\frac{5}{18}} = \frac{4}{5}.$$

A similar reasoning applies if both drew black balls. If a_2 draws an opposite color of ball from a_1, then the probabilities for U_W and U_B become equal. For simplicity we will assume that any individual faced with equal probability for U_W and U_B will guess for the urn that contains more balls of the color she saw herself: if a_1 drew a white ball and a_2 draws a black ball, a_2 will guess U_B.[1] Hence, independent of which ball a_1 draws, a_2 will always guess for the urn matching the color of her privately drawn ball. We assume that this tie-breaking rule is common knowledge among the agents too. In this way, every individual following a_2 can also infer the color of a_2's ball.

When a_3 enters, a cascade can arise. If a_1 and a_2 drew balls of different colors, a_3 is rational to guess for the urn that matches the color of the ball she draws. Nevertheless, if a_1 and a_2 drew the same color of balls (given the reasoning previously described, a_3 will know this), say both white, then no matter what color of ball a_3 draws the posterior probability of U_W will be higher than the probability of U_B (and if a_1 and a_2 both drew black balls the other way around). To check this let us calculate the probability of U_W given that a_1 and a_2 drew white balls and a_3 draws a black one:

$$P(U_W|w_1, w_2, b_3) = \frac{P(U_W) \cdot P(w_1, w_2, b_3|U_W)}{P(w_1, w_2, b_3)} = \frac{\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3}}{\frac{1}{9}} = \frac{2}{3}.$$

It is obvious that $P(U_W|w_3, w_2, w_1)$ will be even larger, thus whatever ball a_3 draws, it will be rational for her to guess for U_W. Hence, if a_1 and a_2 draw the same color of balls a cascade will start from a_3 on![2] The individuals following a_3 should therefore take a_3's guess as conveying no new information. Furthermore, everyone after a_3 will have the same information as a_3 (the information about what a_1 and a_2 drew) and their reasoning will therefore be identical to the one of a_3 and the cascade will continue.

[1] This tie-breaking rule is a simplifying assumption but it does not affect the likelihood of cascades arising. Moreover, there seems to be some empirical evidence that this is what most people do and it is also a natural tie-breaking rule if the individuals assign a small chance to the fact that other people might make errors [6].

[2] Note that the cascade will start even if we change the tie-breaking rule of a_2 such that she randomizes her guess whenever she draws a ball contradicting the guess of a_1. In this case, if a_1 and a_2 guess for the same urn, a_3 will not know the color of a_2's ball, but she will still consider it more likely that a_2's ball matches the ball of a_1 and hence consider it more likely that the urn which they have picked is the one in the room.

If U_B is, as we assumed, the urn actually placed in the room and both a_1 and a_2 draw white balls (which happens with probability $\frac{1}{9}$) then a cascade leading to everyone making the *wrong* guess starts. Note, however, that if both a_1 and a_2 draw black balls (which happens with probability $\frac{4}{9}$), then a cascade still starts, but this time it will lead to everyone making the *right* guess. Thus, when a cascade happens it is four times more likely in this example that it leads to right guesses than to the wrong guesses. This already supports the claim that rational agents can be well aware of the fact that they are in a cascade without it forcing them to change their decisions.

The general conclusion of this example is that even though informational cascades can look irrational from a social perspective, they are not irrational from the perspective of any individual participating in them.

3.2.3 Objections

The above semi-formal analysis summarizes the standard Bayesian treatment of the urn example, as given e.g. in [71]. However, as we mentioned in Section 3.1.3, several objections may be raised against the way the conclusion is reached by this analysis. First of all, the example has only been partially formalized, in the sense that the public announcements of the individuals' guesses are not explicitly present in it, neither is the reasoning that lets the individuals ignore the guesses of the previous people caught in a cascade. Moreover, the proposed analysis does not formally capture the agents' full higher-order reasoning (i.e. their reasoning about the others' beliefs and about the others' higher-order reasoning about their beliefs, etc). Therefore, the above argument does not rule out the possibility that some kind of higher-order reflection may help prevent (or break) an informational cascade: it might be the case that, after realizing that they are participating in a cascade, agents may use this information to try to stop the cascade. For all these reasons, it seems necessary to give a more complete analysis, capturing both the communication and the full higher-order reasoning of the agents. This is precisely what we will do in the next sections: we will check whether the same conclusion (the inescapability of cascades by rational means) is reached when agents are dotted with more reasoning powers.

3.3 Probabilistic DEL Modeling

In this section we will analyze cascades using the tools of Probabilistic Dynamic Epistemic Logic [48, 112]. Our presentation will be based on a simplified version of the framework from [48], in which agents are introspective as far as their own subjective probabilities are concerned (so that an agent's subjective probability assignment does not depend on the actual state of the world but only on that world's partition cell in the agent's information partition). We also use a slightly different graphic representation, which makes explicit the *odds* between any two possible states (considered pairwise) according to each agent. This allows us to present directly a *comparative* treatment of the rational guess of each agent and will make obvious the similarity with the framework for "counting evidences" that we will introduce in Section 3.4. We start with some definitions.

3.3.1 State Models

3.3.1. DEFINITION. [Probabilistic Epistemic State Models] A probabilistic multi-agent epistemic state model $\mathcal{M}$ is a structure $(S, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, (P_a)_{a\in\mathcal{A}}, \Psi, \|\bullet\|)$ such that:

- S is a set of states (or "possible worlds");
- $\mathcal{A}$ is a set of agents;
- for each agent a, $\sim_a \subseteq S \times S$ is an equivalence relation interpreted as agent a's epistemic indistinguishability. This captures the agent's hard information about the actual state of the world;
- for each agent a, $P_a : S \to [0, 1]$ is a map that induces a probability measure on each $\sim_a$-equivalence class (i.e., we have $\sum\{P_a(s') : s' \sim_a s\} = 1$ for each $a \in \mathcal{A}$ and each $s \in S$). This captures the agent's subjective probabilistic information about the state of the world;
- Ψ is a given set of "atomic propositions", denoted by $p, q, \ldots$. Such atoms p are meant to represent *ontic "facts'* that might hold in a world.
- $\|\bullet\| : \Psi \to \mathcal{P}(S)$ is a "valuation" map, assigning to each atomic proposition $p \in \Psi$ some set of states $\|p\| \subseteq S$. Intuitively, the valuation tells us which facts hold in which worlds.

3.3.2. DEFINITION. [Relative Likelihood] The *relative likelihood* (or "odds") of a state s against a state t according to agent a, $[s : t]_a$, is defined as

$$[s : t]_a := \frac{P_a(s)}{P_a(t)}.$$

We will draw probabilistic epistemic state models in the following way (see Figure 3.1 below and the following ones): each state is drawn as an oval, containing the name of the state and the facts p that are "true" at the state (i.e. the atomic sentences p such that their valuation $\|p\|$ contains this state). For each agent $a \in \mathcal{A}$, we draw a-labeled arrows going from each state s towards all the states in the same a-information cell to which a attributes equal or higher odds (than to state s). Therefore, the qualitative arrows represent both the hard information (indistinguishability relation) and the probability ordering relative to an agent: arrows point towards the indistinguishable states that she considers to be at least as probable. To make explicit the odds assigned by agents to states, we label these arrows with the quantitative information (followed by the agents' names in the brackets). For instance, the fact that $[s : t]_a = \frac{\alpha}{\beta}$ is encoded by an a-arrow from state s to state t labeled with the quotient $\alpha : \beta(a)$. For simplicity, we don't represent the loops relating each state to itself, in every model, since every state of every model is trivially a-indistinguishable from itself for each agent, with equal odds $(1 : 1)$ to itself.

The initial situation. To illustrate a probabilistic epistemic state model with odds, consider the initial situation of our urn example presented in Section 3.2 as pictured in Figure 3.1. In this initial model $\mathcal{M}_0$, it is equally probable that U_W or U_B is true (and therefore the prior odds are equal) and all agents know this. The actual state (denoted by the thicker oval) s_B satisfies the proposition U_B, while the state s_W satisfies the proposition U_W. The bidirectional arrow labeled with "1:1 (all a)" represents the fact that all agents consider both states equally probable.

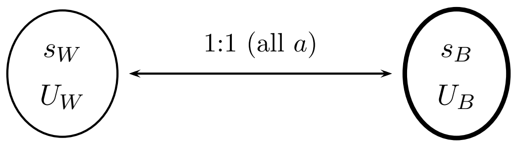


Figure 3.1: The initial probabilistic state model $\mathcal{M}_0$ of the urn example.

3.3.3. Definition. [Epistemic-probabilistic language] As in [48], the "static" language we adopt to describe these models is the epistemic-probabilistic language due to Halpern and Fagin [75]. The syntax is given by the following Backus-Naur form:

$$\varphi \ := \ p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_a\varphi \mid \alpha_1 \cdot P_a(\varphi) + \ldots + \alpha_n \cdot P_a(\varphi) \geq \beta$$

where $p \in \Psi$ are atomic propositions, $a \in \mathcal{A}$ are agents and $\alpha_1, \ldots, \alpha_n, \beta$ stand for arbitrary rational numbers. Let us denote this language by $\mathcal{L}$.

The *semantics* is given by associating to each formula φ and each model $\mathcal{M} = (S, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, (P_a)_{a\in\mathcal{A}})$, some *interpretation* $\|\varphi\|_{\mathcal{M}} \subseteq S$, given recursively by the obvious inductive clauses[3]. If $s \in \|\varphi\|_{\mathcal{M}} \subseteq S$, then we say that φ is *true* at state s (in model $\mathcal{M}$).

[3]It is worth noting that, when checking whether a given state s belongs to $\|\varphi\|$, every expression of the form $P_a(\psi)$ is interpreted conditionally on agent a's knowledge at s, i.e. as $P_a(\|\psi\| \cap \{s' \in S : s' \sim_a s\})$. See [75], [48] for details.

In this language, one can introduce strict inequalities, as well as equalities, as abbreviations, e.g.:

$$P_a(\varphi) > P_a(\psi) \ := \ \neg(P_a(\psi) - P_a(\varphi) \geq 0),$$
$$P_a(\varphi) = P_a(\psi) \ := \ (P_a(\varphi) - P_a(\psi) \geq 0) \wedge (P_a(\psi) - P_a(\varphi) \geq 0)$$

One can also define an expression saying that *an agent a assigns higher odds to φ than to ψ* (given her current information cell):

$$[\varphi : \psi]_a > 1 \ := \ P_a(\varphi) > P_a(\psi)$$

3.3.2 Event Models

To model the incoming of new information, we use *probabilistic event models*, as introduced by van Benthem et al. [48]: these are a probabilistic refinement of the notion of *event models* of Dynamic Epistemic Logic [24]. Here we use a simplified setting, which assumes introspection of subjective probabilities.

3.3.4. Definition. [Probabilistic Event Models] A probabilistic event model $\mathcal{E}$ is a sextuple $(E, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, (P_a)_{a\in\mathcal{A}}, \Phi, pre)$ such that:

- E is a set of possible events,
- $\mathcal{A}$ is a set of agents;
- $\sim_a \subseteq E \times E$ is an equivalence relation interpreted as agent a's epistemic indistinguishability between possible events, capturing a's hard information about the event that is currently happening;
- P_a gives a probability assignment for each agent a and each $\sim_a$-information cell. This captures some new, independent subjective probabilistic information gained by the agent during the event: when observing the current event (without using any prior information), agent a assigns probability $P_a(e)$ to the possibility that in fact e is the actual event that is currently occurring;
- Φ is a finite set of mutually inconsistent propositions (in the above probabilistic-epistemic language $\mathcal{L}$), called *preconditions*;
- *pre* assigns a probability distribution $pre(\bullet|\phi)$ over E for every proposition $\phi \in \Phi$. This is an "occurrence probability": $pre(e|\phi)$ expresses the prior probability that event $e \in E$ might occur in a(ny) state satisfying precondition ϕ.

As before, the probability P_a can alternatively be expressed as probabilistic odds $[e : e']_a$ for any two events e, e' and any agent a. Our event models are drawn in the same fashion as our state models above: for each agent a, a-arrows go from a possible event e towards all the events (of a's information cell) to which a attributes equal or higher odds.

The first agent's observation. As an example of an event model, consider the first observation of a ball in our urn case, as represented in the model $\mathcal{E}_1$ from Figure 3.2. Here a_1 draws a white ball from the urn and looks at it. According to all the other agents, two events can happen: either a_1 observes a white ball (the actual event w_1, denoted by a thicker frame) or she observes a black one (event b_1). Moreover, only agent a_1 knows which event is the actual one. The expressions $pre(U_W) = \frac{2}{3}$ and $pre(U_B) = \frac{1}{3}$ depicted at event w_1 represents that the prior probabilities $pre(w_1 \mid U_W)$ that event w_1 occurs when U_W is satisfied is $\frac{2}{3}$ while the probability $pre(w_1 \mid U_B)$ that event w_1 happens when U_B is satisfied is $\frac{1}{3}$ (and vice versa for event b_1). The bidirectional arrow for all agents except a_1 represents the fact that agent a_1 can distinguish between the two possible events (since she *knows* that she sees a white ball), while the others cannot distinguish them and have (for now) no reason to consider one event more likely than the other, i.e., their odds are 1 : 1.

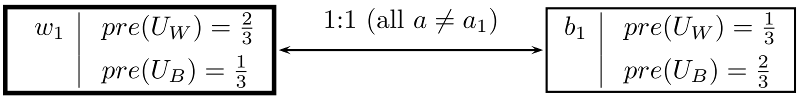


Figure 3.2: The probabilistic event model $\mathcal{E}_1$ of agent a_1 drawing a white ball.

3.3.3 Product Update

To model the change of odds after new information is received, we now combine probabilistic epistemic state models with probabilistic event models using a notion of product update.

3.3.5. Definition. [Probabilistic Product Update] Given a probabilistic epistemic state model $\mathcal{M} = (S, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, (P_a)_{a\in\mathcal{A}}, \Psi, \|\bullet\|)$ and a probabilistic event model $\mathcal{E} = (E, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, (P_a)_{a\in\mathcal{A}}, \Phi, pre)$, the updated state model $\mathcal{M} \otimes \mathcal{E} = (S', \mathcal{A}, (\sim'_a)_{a\in\mathcal{A}}, (P'_a)_{a\in\mathcal{A}}, \Psi', \| \bullet \|')$, is given by:

$$S' = \{(s, e) \in S \times E \mid pre(e \mid s) \neq 0\},$$

$$\Psi' = \Psi,$$

$$\|p\|' = \{(s, e) \in S' : s \in \|p\|\},$$

$$(s, e) \sim'_a (t, f) \text{ iff } \; s \sim_a t \text{ and } e \sim_a f,$$

$$P'_a(s, e) = \frac{P_a(s) \cdot P_a(e) \cdot pre(e \mid s)}{\sum\{P_a(t) \cdot P_a(f) \cdot pre(f \mid t) : s \sim_a t, e \sim_a f\}},$$

where we used the notation

$$pre(e \mid s) := \sum \{pre(e \mid \phi) : \phi \in \Phi \text{ such that } s \in \|\phi\|_{\mathcal{M}}\}$$

(so that $pre(e \mid s)$ is either $= pre(e|\phi_s)$ where ϕ_s is the unique precondition in Φ such that ϕ_s is true at s, or otherwise $pre(e \mid s) = 0$ if no such precondition ϕ_s exists).

This definition can be justified on Bayesian grounds: the definition of the new indistinguishability relation simply says that the agent puts together her old and new hard information[4]; while the definition of the new subjective probabilities is obtained by multiplying the old probability previously assigned to event e (obtained by applying the conditioning rule $P_a(e) = P_a(s) \cdot P_a(e \mid s) = P_a(s) \cdot pre(e \mid \phi_s)$) with the new probability independently assigned (without using any prior information) to event e during the event's occurrence, and then renormalizing to incorporate the new hard information. The reason for using multiplication is that the two probabilities of e are supposed to represent two independent pieces of probabilistic information.[5]

Again, it is possible, and even easier, to express posterior probabilities in terms of posterior relative likelihoods:

$$[(s,e) : (t,f)]_a = [s : t]_a \cdot [e : f]_a \cdot \frac{pre(e \mid s)}{pre(f \mid t)}.$$

Result of the first agent's observation. The result of the product update of the initial state model $\mathcal{M}_0$ from Fig. 3.1 with the event model $\mathcal{E}_1$ of Fig. 3.2 is given by the new model $\mathcal{M}_0 \otimes \mathcal{E}_1$ of Fig. 3.3. The upper right state is the actual situation, in which U_B is true, but in which the first ball which has been observed was a white one. Agent a_1 *knows* that she observed a white ball (w_1), but she does not know which urn is the actual one, so her actual information cell consists of the upper two states, in which she considers U_W to be twice as likely as U_B. The other agents still cannot exclude any possibility. This is going to change once the first agent announces her guess.

[4]This is the essence of the "Product Update" introduced by Baltag et alia [24], which forms the basis of most widespread versions of Dynamic Epistemic Logic.

[5]In fact, this feature is irrelevant for our analysis of cascades: no new non-trivial probabilistic information is gained by the agents during the events forming our cascade example. This is reflected in the fact that, in our analysis of cascades, we will use only event models in which the odds are 1 : 1 between any two indistinguishable events.

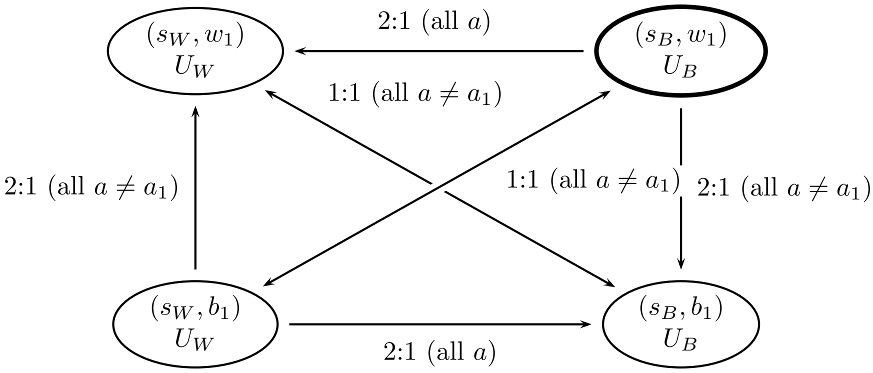


Figure 3.3: The updated probabilistic state model $\mathcal{M}_0 \otimes \mathcal{E}_1$ after a_1 draws a white ball.

3.3.4 Public Announcement

To model the agents' announcements of their guesses, we will use the standard *public announcements* of [133], where a (truthful) public announcement $!\varphi$ of a proposition φ is an event which has the effect of deleting all worlds of the initial state model that do not satisfy φ. Note that, public announcements $!\varphi$ can be defined as a special kind of probabilistic event models: take $E = \{e_{!\varphi}\}$, $\sim_a = \{(e_{!\varphi}, e_{!\varphi})\}$, $\Phi = \{\varphi\}$, $pre(e_{!\varphi} \mid \varphi) = 1$, $P_a(e_{!\varphi}) = 1$.

The first agent announces her guess. Now, after her private observation, agent a_1 publicly announces that she considers U_W to be more likely than U_B. This is a public announcement $!([U_W : U_B]_{a_1} > 1)$ of the sentence $[U_W : U_B]_{a_1} > 1$ (as defined above as an abbreviation in our language), expressing the fact that agent a_1 assigns higher odds to urn U_W than to urn U_B. Since all agents know that the only reason a_1 could consider U_W more likely than U_B is that she drew a white ball (her announcement can be truthful only in the situations in which she drew a white ball), the result is that all agents come to know this fact. This is captured by our modelling, where her announcement simply erases the states (s_W, b_1) and (s_B, b_1) and results in the new model $\mathcal{M}_1$ of Fig. 3.4.

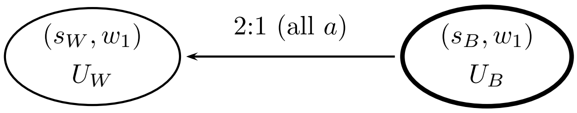


Figure 3.4: The updated probabilistic state model $\mathcal{M}_1$ after a_1's announcement.

The second agent. By repeating the above reasoning, we know that, after another observation of a white ball by agent a_2 (the event model is as above in Fig. 3.2 but relative to agent a_2 instead of agent a_1) and a similar public announcement of $[U_W : U_B]_{a_2} > 1$, the resulting state model $\mathcal{M}_2$, depicted in Fig. 3.5, will be such that all agents now consider U_W *four* times more likely than U_B.

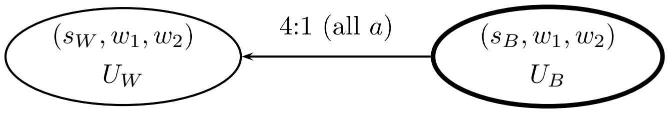


Figure 3.5: The updated probabilistic state model $\mathcal{M}_2$ after a_2's announcement.

3.3.5 The Birth of a Cascade

The third agent. Let us now assume that agent a_3 enters the room and privately observes a black ball. The event model $\mathcal{E}_3$ of this action is in Figure 3.6, and is again similar to the earlier event model (Fig. 3.2) but relative to agent a_3 and this time, since a *black* ball is observed, the actual event is b_3.

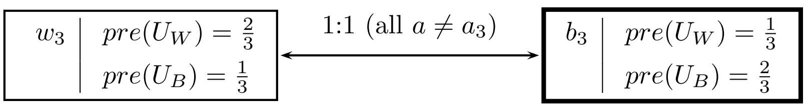


Figure 3.6: The probabilistic event model $\mathcal{E}_3$ of a_3 drawing a black ball.

The result of a_3's observation is then given by the updated state model $\mathcal{M}_2 \otimes \mathcal{E}_3$ shown in Figure 3.7.

Since only agent a_3 knows what she has observed, her actual information cell only contains the states in which the event b_3 has happened, while all other agents cannot distinguish between the four possible situations. Moreover, agent a_3 still considers U_W more probable than U_B, *irrespective* of the result of her private observation (w_3 or b_3). So the fact that $[U_W : U_B]_{a_3} > 1$ is now *common knowledge* (since it is true at all states of the entire model). This means that announcing this fact, via a new public announcement of $[U_W : U_B]_{a_3} > 1$ will *not* delete any state: the model $\mathcal{M}_3$ after the announcement is simply *the same as before* (Fig. 3.7).

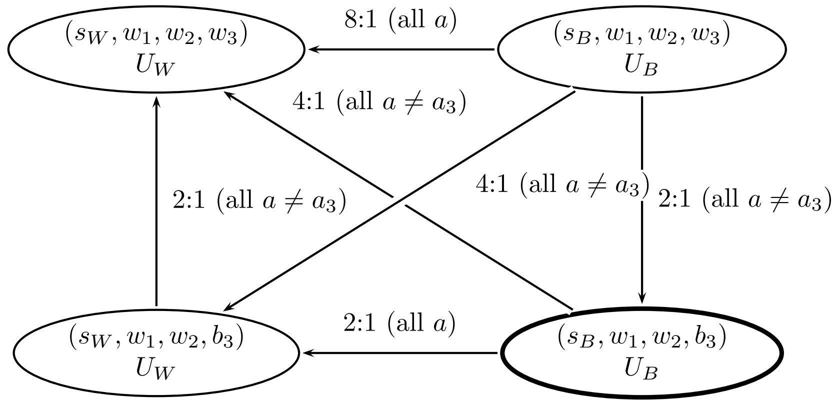


Figure 3.7: The probabilistic state model $\mathcal{M}_2 \otimes \mathcal{E}_3$ after a_3 draws a black ball,

So the *third agent's public announcement bears no information whatsoever*: an informational cascade has been formed, even though all agents have reasoned correctly about probabilities. From now on, the situation will keep repeating itself: although the state model will keep growing, *all agents will always consider* U_W *more probable than* U_B *in all states* (irrespective of their own observations). This is shown formally by the following result.

3.3.6. PROPOSITION. *Starting in the model in Fig. 3.1 and following the above protocol, we have that: after* $n-1$ *private observations and public announcements* $e_1, !([U_W : U_B]_{a_1} > 1) \ldots, e_{n-1}, !([U_W : U_B]_{a_{n-1}} > 1)$ *by agents* $a_1, \ldots, a_{n-1}$*, with* $n \geq 3$*,* $e_1 = w_1$ *and* $e_2 = w_2$*, the new state model* $\mathcal{M}_{n-1}$ *will satisfy*

$$[U_W : U_B]_a > 1, \text{ for all } a \in \mathcal{A}.$$

To show this, we prove the following stronger proposition:

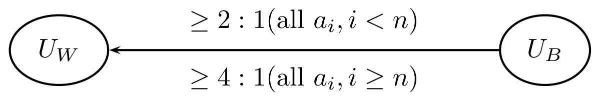
3.3.7. PROPOSITION. *After* $n-1$ *private observations and announcements as above, the new state model* $\mathcal{M}_{n-1}$ *will satisfy*

$$[U_W : U_B]_{a_i} \geq 2, \text{ for all } i < n, \text{ and}$$

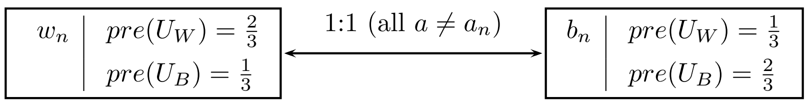
$$[U_W : U_B]_{a_i} \geq 4, \text{ for all } i \geq n.$$

From this claim, the desired conclusion follows immediately.

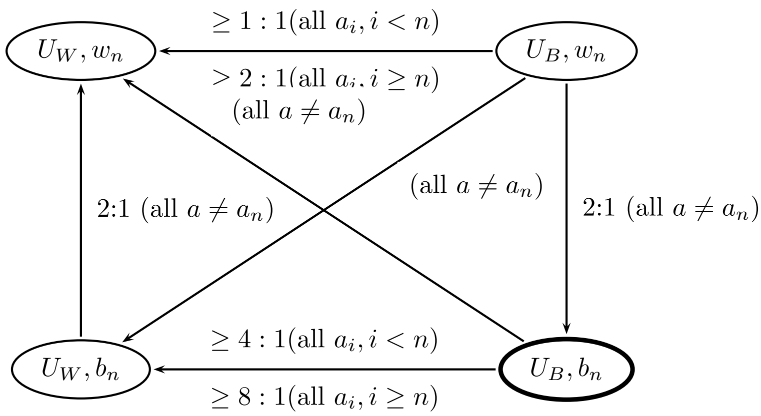
Proof: We give only a sketch of the proof, using an argument based on partial descriptions of our models. The base case $n = 3$ was already proved above. Assume the inductive hypothesis for $n-1$. By lumping together all the U_W-states in $\mathcal{M}_{n-1}$, and similarly all the U_B-states, we can represent this hypothesis via the following partial representation of $\mathcal{M}_{n-1}$:



Note that this is just a "bird's view" representation: the actual model $\mathcal{M}_{n-1}$ has 2^{n-2} states. To see what happens after one more observation e_n by agent n, take the update produce of this representation with the event model $\mathcal{E}_n$, given by:

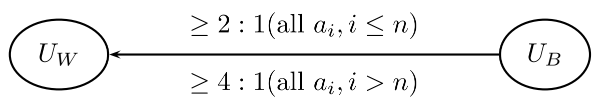


The resulting product is:



where for easier reading we skipped the numbers representing the probabilistic information associated to the diagonal arrows (numbers which are not relevant for the proof).

By lumping again together all indistinguishable U_W-states in $\mathcal{M}_{n-1}$, and similarly all the U_B-states, and reasoning by cases for agent a_n (depending on her actual observation), we obtain:



Again, this is just a bird's view: the actual model has 2^n states. But the above partial representation is enough to show that, in this model, we have $[U_W : U_B]_{a_i} \geq 2$ for all $i < n + 1$, and $[U_W : U_B]_{a_i} \geq 4$ for all $i \geq n + 1$.

Since in particular $[U_W : U_B]_{a_n} > 1$ holds in all the states, this fact is common knowledge: so, after publicly announcing it, the model stays the same! Hence, we proved the induction step for n. □

So, in the end, all the guesses will be wrong: the whole group will assign a higher probability to the wrong urn (U_W). Thus, we have proved that *individual Bayesian rationality with perfect higher-level reflective powers can still lead to "group irrationality"*. This shows that in some situations there simply is no higher-order information available to any of the agents to prevent them from entering the cascade; not even the information that they *are* in a cascade can help in this case. (Indeed, in our model, after the two guesses for U_W of a_1 and a_2, it is already common knowledge that a cascade has been formed!)

3.3.6 Objection

A possible objection to the model presented in this section is that it still relies on the key assumption that the involved agents are perfect *Bayesian* reasoners. However, many authors argue that *rationality* cannot be identified with *Bayesian rationality*. There are other ways of reasoning that can be deemed rational without involving doing cumbersome Bayesian calculations. It is therefore possible to object to our formalization of rational cascades which lead a group to collective failure that it relies on this unrealistic notion of rationality. In practice, many people seem to use much simpler *"counting" heuristics*, e.g. guessing U_W when one has more pieces of evidence in favor of U_W than in favor of U_B, i.e. one knows that more white balls were drawn than black balls. Hence, the next section will turn to a model of informational cascades based on simple *counting* instead of Bayesian updates.

3.4 Counting DEL Modeling

To address the above objection, according to which an agent can be taken to be rational without having to be a perfect probabilistic reasoner, this section presents a formalized setting of the same urn example using a notion of rationality based on a simple counting heuristic instead of full-fledge Bayesian probabilities. The logical framework for this purpose is inspired by the probabilistic framework of the previous section. However, it is substantially simpler: instead of calculating the probability of a given possible state, agents simply count the evidence in favor of this state. In a nutshell, an agent is deemed rational as long as he can count and compare two numbers to decide which one is the biggest.

3.4.1 State Models

More precisely, we label each state with a number representing the strength of all evidence in favor of that state being the actual one. This intuition is represented in the following formal definition:

3.4.1. Definition. [Counting Epistemic Models] A counting multi-agent epistemic model $\mathcal{M}$ is a structure $(S, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, f, \Psi, \| \bullet \|)$ such that:

- S is a set of states,
- $\mathcal{A}$ is a set of agents,
- $\sim_a \subseteq S \times S$ is an equivalence relation interpreted as agent a's epistemic indistinguishability,
- $f : S \to \mathbb{N}$ is an "evidence-counting" function, assigning a natural number to each state in S,
- Ψ is a given set of atomic sentences,
- $\| \bullet \| : \Psi \to \mathcal{P}(S)$ is a valuation map.

The initial situation. We can now represent the initial situation of the urn example by the model of Figure 3.8. The two possible states s_W and s_B correspond to U_W (resp. U_B) being placed in the room. The notation U_W ; 0 at the state s_W represents that $f(s_W) = 0$ and that the atomic proposition U_W is true at s_W (and all other atomic propositions are false). The line between s_W and s_B labeled by "all a" means that the two states are indistinguishable for all agents a. Finally, the thicker line around s_B represents that s_B is the actual state.

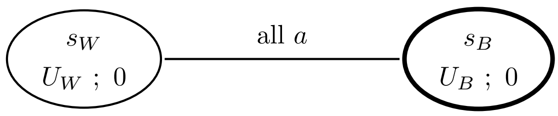


Figure 3.8: The initial counting model of the urn example.

3.4.2 Event Models

We now turn to how to update counting epistemic models. However, first note that, at this stage, there is not much that distinguishes counting epistemic models from probabilistic ones. In the case where models are finite, one can simply sum up the values of $f(w)$ for all states w in a given information cell and rescale $f(w)$ by this factor thereby obtaining a probabilistic model from a counting model. Additionally, assuming that all probabilities are rational numbers one can easily move the other way as well. Despite this similarity, when we address dynamic issues, the counting framework becomes significantly simpler than the probabilistic one. Indeed, we will not need to use multiplication together with Bayes' rule and renormalization, we can simply use *addition*. More formally:

3.4.2. Definition. [Counting Event Models] A counting event model $\mathcal{E}$ is a quintuple $(E, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, \Phi, pre)$ such that:

- E is a set of possible events,
- $\mathcal{A}$ is a set of agents,
- $\sim_a \subseteq E \times E$ is an equivalence relation interpreted as agent a's epistemic indistinguishability,
- Φ is a finite set of pairwise inconsistent propositions,
- $pre : E \to (\Phi \to (\mathbb{N} \cup \{\perp\}))$ is a function from E to functions from Φ to the natural numbers (extended with $\perp$)[6]. It assigns to each event $e \in E$ a function $pre(e)$, which to each proposition $\phi \in \Phi$ assigns the strength of evidence that the event e provides for ϕ.

The first agent's observation. As an example of a counting event model, the event model of the first agent drawing a white ball is shown in Figure 3.9. In this event model there are two events w_1 and b_1, where the actual event is w_1 (marked by the thick box). A notation like $pre(U_W) = 1$ at w_1 simply means that $pre(w_1)(U_W) = 1$.[7] Finally, the line between w_1 and b_1 labeled "all $a \neq a_1$" represents that the events w_1 and b_1 are indistinguishable for all agents a except a_1.

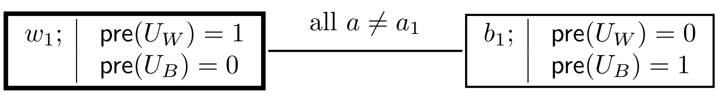


Figure 3.9: The counting event model of a_1 drawing a white ball.

[6]Here, $\perp$ essentially means "undefined": so it is just an auxiliary symbol used to describe the case when *pre* is a *partial* function.

[7]To fit the definition of counting event models properly, U_W and U_B must be pairwise inconsistent, however, this claim fits perfectly with the example where only one of the urns is placed in the room and we could simple replace U_W by $\neg U_B$.

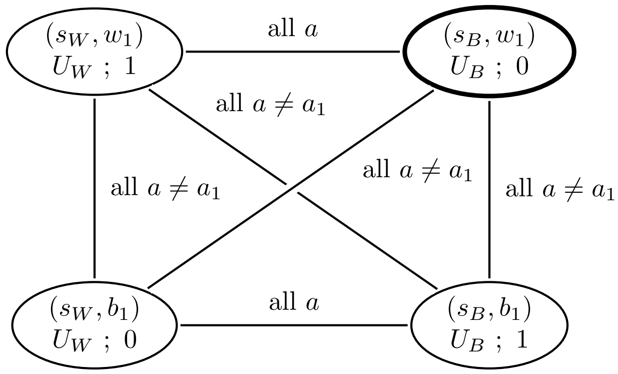


Figure 3.10: The updated counting model after a_1 draws a white ball.

3.4.3 Product Update

A counting epistemic model is updated with a counting event model in the following way:

3.4.3. Definition. [Counting Product Update]

Given a counting epistemic model $\mathcal{M} = (S, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, f, \Psi, \|\bullet\|)$ and a counting event model $\mathcal{E} = (E, \mathcal{A}, (\sim_a)_{a\in\mathcal{A}}, pre)$, we define the product update $\mathcal{M} \otimes \mathcal{E} = (S', \mathcal{A}, (\sim'_a)_{a\in\mathcal{A}}, f', \Psi', \|\bullet\|)$ by

$$S' = \{(s, e) \in S \times E \mid pre(s, e) \neq \bot\},$$

$$\Psi' = \Psi,$$

$$\|p\|' = \{(s, e) \in S' : s \in \|p\|\},$$

$$(s, e) \sim_a (t, f) \text{ iff } s \sim_a t \text{ and } e \sim_a f,$$

$$f'((s, e)) = f(s) + pre(s, e), \quad \text{for } (s, e) \in S',$$

where we used the notation $pre(s, e)$ to denote $pre(e)(\phi_s)$ for the unique $\phi_s \in \Phi$ such that $s \in \|\phi_s\|_{\mathcal{M}}$, if such a precondition $\phi_s \in \Phi$ exist, and otherwise we put $pre(s, e) = \bot$.

The result of the first agent's observation. With this definition we can now calculate the product update of the models of the initial situation (Fig. 3.8) and the first agent drawing a white ball (Fig. 3.9). The resulting model is shown in Figure 3.10.

3.4.4 Public Announcement

We need to say how we will represent the action that agent a_1 guesses urn U_W. As in the case of probabilistic modeling we will interpret this action as a public announcement. A public announcement of ϕ in the classical sense of eliminating all non-ϕ states, is a special case of a counting event model with just one event e, $\Phi = \{\phi\}$, $\sim_a = \{(e,e)\}$ for all $a \in \mathcal{A}$, and $pre(e)(\phi) = 0$. Setting $pre(e)(\phi) = 0$ reflects the choice that we take public announcements not to provide any increase in the strength of evidence for any possible state, but only revealing hard information about which states are possible. In the urn example it is the drawing of a ball from the urn that increases the strength of evidence, whereas the guess simply conveys information about the announcer's hard information about the available evidence for either U_W or U_B. As in the previous section, we will interpret the announcements as revealing whether their strength of evidence for U_W is smaller or larger than their strength of evidence for U_B.

We therefore require a formal language that contains formulas of the form $\phi <_a \psi$, for all formulas ϕ and ψ. The semantics of the new formula is given by:

$$\|\phi <_a \psi\|_{\mathcal{M}} = \{s \in S \mid f(a, s, \|\phi\|_{\mathcal{M}}) < f(a, s, \|\psi\|_{\mathcal{M}})\},$$

where for any given counting model $\mathcal{M} = (S, (\sim_a)_{a\in\mathcal{A}}, f, \Psi, \| \bullet \|)$ and any set of states $T \subseteq S$ we use the notation

$$f(a, s, T) := \sum \{f(t) : t \in T \text{ such that } \ t \sim_a s\}.$$

The first agent announces her guess. The event that agent a_1 announces that she guesses in favor of U_W will be interpreted as a public announcement of $U_B <_{a_1} U_W$. This proposition is only true at the states (s_W, w_1) and (s_B, w_1) of the above model and thus the states (s_W, b_1) and (s_B, b_1) are removed in the resulting model shown in Figure 3.11.

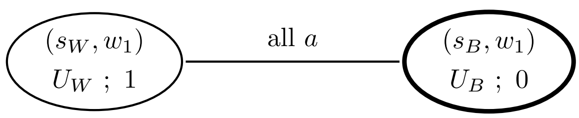


Figure 3.11: The counting model after a_1 publicly announces that $U_B <_{a_1} U_W$.

The second agent. Moreover, the event that a_2 draws a white ball can be represented by an event model identical to the one for agent a_1 drawing a white ball (Fig. 3.9) except that the label on the drawn relation should be changed to "all $a \neq a_2$". The updated model after the event that a_2 draws a white ball will look as shown in Figure 3.12. Note that in this updated model, $U_B <_{a_2} U_W$ is only true at (s_W, w_1, w_2) and (s_B, w_1, w_2), thus when a_2 announces her guess for U_W (interpreted as a public announcement of $U_B <_{a_2} U_W$) the resulting model is pictured in Figure 3.13.

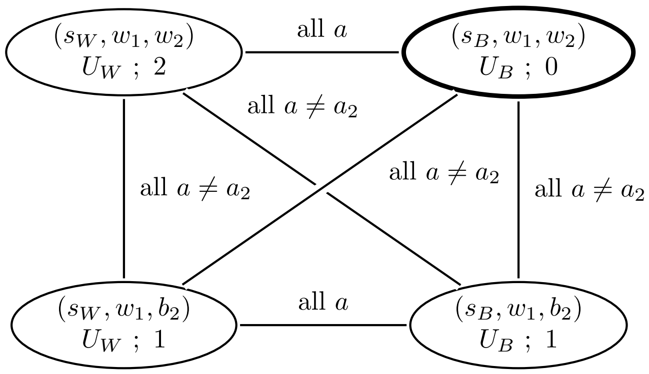


Figure 3.12: The updated counting model after a_2 draws a white ball.

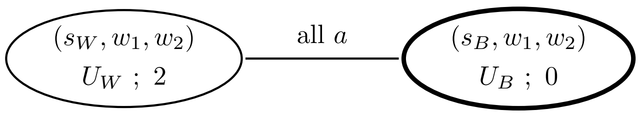


Figure 3.13: The counting model after a_2 publicly announces that $U_B <_{a_2} U_W$.

3.4.5 The Birth of a Cascade

The third agent. Assuming that agent a_3 draws a black ball this can be represented by an event model almost identical to the one for agent a_1 drawing a white ball (Fig. 3.9). The only differences are that the label on the line should be changed to "all $a \neq a_3$)" and the actual event should be b_3. Updating the model of Figure 3.13 with this event will result in the model of Figure 3.14.

Note that in Figure 3.14 the proposition $U_B <_{a_3} U_W$ is true in the entire model. Hence, agent a_3 has more evidence for U_W than U_B and thus, no states will be removed from the model when she announces her guess for U_W (a public announcement of $U_B <_{a_3} U_W$). If a_3 had drawn a white ball instead, the only

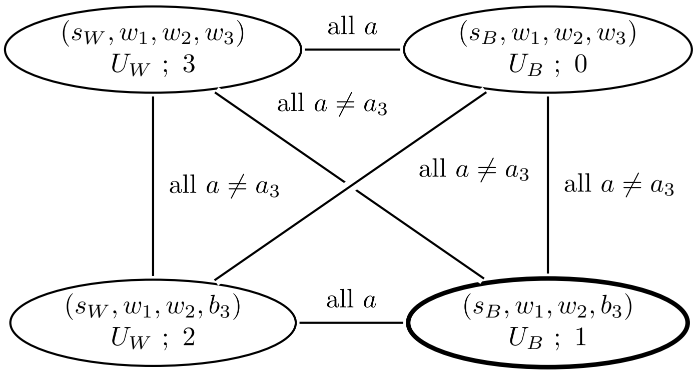


Figure 3.14: The updated counting model after a_3 draws a black ball.

thing that would have be different in the model of Figure 3.14 is that the actual state would be (s_B, w_1, w_2, w_3). Therefore, this would not change the fact that a_3 guesses for U_W and this announcement will remove no states from the model either. In this way, none of the following agents gain any information from learning that a_3 guessed for U_W.

Subsequently whenever an agent draws a ball, she will have more evidence for U_W than for U_B. Thus, the agents will keep guessing for U_W. However, these guesses will not delete any more states. Hence, the models will keep growing exponentially reflecting the fact that no new information is revealed. In other words, *an informational cascade has started.* Formally, one can show the following result:

3.4.4. Proposition. *Let $\mathcal{M}_n$ be the updated model after agent a_n draws either a white or a black ball. Then, if both a_1 and a_2 draw white balls (i.e. we are in the model of Fig. 3.12), then for all $n \geq 3$, $U_B <_{a_n} U_W$ will be true in all states of $\mathcal{M}_n$.*

In words: after the first two agents have drawn white balls all the following agents will all have more evidence for U_W than U_B (no matter which color of ball they draw) and will therefore guess for U_W, however, these guesses will be uninformative to the subsequent agents as the public announcement of $U_B <_a U_W$ will delete no worlds.

Before we can prove this proposition we need some definitions. For all $n \geq 3$, let $\mathcal{E}_n$ be the event model expressing that agent a_n draws either a white ball (w_n) or a black ball (b_n)[8], for instance $\mathcal{E}_1$ is shown in Figure 3.9. Furthermore, let $\mathcal{M}_n$ denote the model obtained after updating with the event $\mathcal{E}_n$, hence $\mathcal{M}_n = \mathcal{M}_{n-1} \otimes \mathcal{E}_n$. The model $\mathcal{M}_3$ is shown in Figure 3.14. We will denote the domain of $\mathcal{M}_n$ by

[8]Which color of ball a_n draws does not matter as it only affects which state will be the actual state.

$dom(\mathcal{M}_n)$. For a proposition ϕ, we will by $f^n(\phi)$ denote $\sum\{f(s) \mid s \in \|\phi\|_{\mathcal{M}_n}\}$. Note that $f^2(U_W) = 2$, $f^3(U_W) = 5$, $f^2(U_B) = 0$, and $f^3(U_B) = 1$.

Now Proposition 2 follows from the following lemma:

3.4.5. LEMMA. *For all $n \geq 3$ the following hold:*

(i) Let $[w]_n := dom(\mathcal{M}_{n-1}) \times \{w_n\}$ and $[b]_n := dom(\mathcal{M}_{n-1}) \times \{b_n\}$. Then $[w]_n$ and $[b]_n$ are the only two information cells of agent a_n in $\mathcal{M}_n$, $dom(\mathcal{M}_n) = ([w]_n \cup [b]_n)$, and $|[w]_n| = |[b]_n| = 2^{n-2}$. Additionally, for all $k > n$, $\mathcal{M}_n$ contains only one information cell for agent a_k, namely the entire $dom(\mathcal{M}_n)$. Furthermore, U_W is true in 2^{n-3} states of $[w]_n$ and 2^{n-3} states of $[b]_n$, and similar, U_B is true in 2^{n-3} states of $[w]_n$ and 2^{n-3} states of $[b]_n$.

(ii) For $s \in [w]_n$, $f(a_n, s, U_W) = f^{n-1}(U_W) + 2^{n-3}$ and $f(a_n, s, U_B) = f^{n-1}(U_B)$. For $s \in [b]_n$, $f(a_n, s, U_W) = f^{n-1}(U_W)$ and $f(a_n, s, U_B) = f^{n-1}(U_B) + 2^{n-3}$.

(iii) $f^n(U_B) + 2^{n-2} < f^n(U_W)$.

(iv) $U_B <_{a_n} U_W$ is true at all states of $\mathcal{M}_n$.

Proof: The proof goes by induction on n. For $n = 3$ the statements $(i) - (iv)$ are easily seen to be true by inspecting the model $\mathcal{M}_3$ as shown in Figure 3.14 of section 3.4. We prove the induction step separately for each of the statements $(i) - (iv)$.

(i): Assume that (i) is true for n. Then, for agent a_{n+1} the model $\mathcal{M}_n$ consists of a single information cell with 2^{n-1} states where U_W is true in half of them and U_B in half of them. Considering the event model $\mathcal{E}_{n+1}$ it is easy to see that updating with this will result in the model $\mathcal{M}_{n+1}$, where there will be two information cells for agent a_{n+1} corresponding to the events w_{n+1} and b_{n+1}, i.e. $[w]_{n+1}$ and $[b]_{n+1}$, and each of these will have 2^{n-1} states. It is also easy to see that for all $k > n+1$ there will only be one information cell for a_k. Finally, it is also easy to see that U_W will be true in 2^{n-2} states of $[w]_n$ and 2^{n-2} states of $[b]_n$ since U_w where true in 2^{n-2} states of $\mathcal{M}_n$ and likewise for U_B.

(ii): Assume that (ii) is true for n. Assume that $s \in [w]_{n+1}$. Then using (i), the definition of the product update, and (i) again, we get:

$$
\begin{aligned}
f(a_{n+1}, s, U_W) &= \sum\{f(t) \mid t \in [w]_{n+1} \cap U_W\} \\
&= \sum\{f((t, w_{n+1})) \mid t \in \mathcal{M}_n, t \in U_W \text{ (in } \mathcal{M}_n)\} \\
&= \sum\{f(t) + pre(w_{n+1})(U_W) \mid t \in \mathcal{M}_n, t \in U_W \text{ (in } \mathcal{M}_n)\} \\
&= \sum\{f(t) + 1 \mid t \in \mathcal{M}_n, t \in U_W \text{ (in } \mathcal{M}_n)\} \\
&= f^n(U_W) + 2^{n-3}.
\end{aligned}
$$

Similarly for U_B we get

$$\begin{aligned} f(a_{n+1}, s, U_B) &= \sum\{f(t) \mid t \in [w]_{n+1} \cap U_B\} \\ &= \sum\{f(t) + pre(w_{n+1})(U_B) \mid t \in \mathcal{M}_n, t \in U_B \text{ (in } \mathcal{M}_n)\} \\ &= \sum\{f(t) + 0 \mid t \in \mathcal{M}_n, t \in U_B \text{ (in } \mathcal{M}_n)\} \\ &= f^n(U_B). \end{aligned}$$

If $s \in [b]_{n+1}$ then,

- $f(a_{n+1}, s, U_W) = f^n(U_W)$ and
- $f(a_{n+1}, s, U_B) = f^n(U_B) + 2^{n-3}$,

follow by reasoning in similar manner.

(*iii*): Assume that $f^n(U_B) + 2^{n-2} < f^n(U_W)$. Note that from (*i*) and (*ii*) we have that

- $f^{n+1}(U_W) = 2f^n(U_W) + 2^{n-2}$,
- $f^{n+1}(U_B) = 2f^n(U_B) + 2^{n-2}$.

But, then

$$\begin{aligned} f^{n+1}(U_B) + 2^{n-1} &= 2f^n(U_B) + 2^{n-2} + 2^{n-1} \\ &= 2(f^n(U_B) + 2^{n-2}) + 2^{n-2} \\ &< 2f^n(U_W) + 2^{n-2} \\ &= f^{n+1}(U_W). \end{aligned}$$

(*v*): Now assume that $U_B <_{a_n} U_W$ is true at all states of $\mathcal{M}_n$. Consider agent a_{n+1}, we then want to prove that $U_B <_{a_{n+1}} U_W$ is true at all states of $\mathcal{M}_{n+1}$. That is, we need to prove that

$$f(a_{n+1}, s, U_B) < f(a_{n+1}, s, U_W),$$

for all $s \in dom(\mathcal{M}_{n+1})$. By (*i*) and the definition of f, we only have to consider two cases, namely when $s \in [w]_{n+1}$ and when $s \in [b]_{n+1}$. Moreover, by (*ii*) we just need to prove that

a) $f^n(U_B) < f^n(U_W) + 2^{n-2}$;

b) $f^n(U_B) + 2^{n-2} < f^n(U_W)$.

Note that *a*) follows from *b*) and *b*) follows directly from (*iii*).

This completes the proof. □

For a full syntactic encoding of an informational cascade within the logical setting, the language has to be changed/extended. Note that the above logical language does not contain a common knowledge operator, which is necessary to render a full syntactic encoding. A solution to this problem is offered in a new framework in [2], which uses a setting that combines a variant of the Logic of Communication and Change from [47] and a variant of Dynamic Epistemic Probabilistic Logic from [48].

3.5 Conclusion

3.5.1 Diagnosis: Too Much (or Not Enough) Information!

Imagine that we change the protocol to forbid all communication. This would amount to making individual guesses entirely private. By Condorcet's Jury Theorem [66][9], we know that in this case, by taking a poll at the end of the protocol, the majority vote would match the correct urn with very high probability, converging to 1 as the number of agents increases to infinity.

This shows that examples such as the one analyzed in this chapter are indeed cases of "epistemic Tragedies of the Commons": situations in which *communication is an obstacle to group truth-tracking*. In these cases, a cascade can be stopped only in two ways: either by "irrational" actions by some of the in-group agents themselves, or else by outside intervention by an external agent with different information or different interests than the group. An example of the first solution is if some of the agents *simply disregard* the information obtained by public communication and make their guess solely on the basis of their own observations: in this way, they lower the probability that their guess is correct (which is "irrational" from their own individual perspective), but they highly increase the probability that the majority guess will be correct. An example of the second solution is if the protocol is modified (or only disrupted) by some external agent with regulative powers (the "referee" or the "government"). Such a referee can simply *forbid communication* (thus returning to the protocol in the Condorcet's Jury Theorem, which assumes independence of opinions). Or she might require *more communication*; e.g. require that the agents should announce, not only their beliefs about the urns, but also their *reasons* for these beliefs: *the evidence* supporting their beliefs. This evidence might be the number of pieces of evidence in favor of each alternative (in the case that they used the counting heuristics); or it might be the subjective probability that they assign to each alternative; or finally, it might be all their available evidence: i.e. the actual color of the ball that

[9]The Condorcet protocol as well as other variations have been fully formalized as strategies in a game theoretic setting in [1].

they observed (since all the rest of their evidence is already public). Requiring agents to share any of these forms of evidence is enough to stop the cascade in the above example.

One may thus argue that *partial communication* (sharing opinions and beliefs, but not sharing the underlying reasons is evidence) is the cause of the collective failure in informational cascades. More (and better) communication, more true deliberation based on sharing arguments and justifications (rather than simple beliefs), *may* sometimes stop the cascade. A "total communication", in which everybody shares *all* their evidence, all their reasons, all the relevant facts, will be an effective way of stopping cascades (provided that no agent lies and that agents perfectly trust each other). In our toy example, this can be easily done: the relevant pieces of evidence are very few. But it is unrealistic to require such total communication in a real-life situation: the number of facts that might be of relevance is practically unlimited, and moreover it might not be clear to the agents themselves which facts are relevant and which not. So in practice this would amount to asking agents to publicly share all their life experiences. With such a protocol, deliberation would never end, and the moment of decision would always be indefinitely postponed.

Therefore, in practice, the danger remains: no matter how rational the agents are, how well-justified their beliefs are, how open they are to communication, how much time they spend sharing their arguments and presenting their evidence, there is always the possibility that all this rational deliberation will only lead the group into a cascading dead-end, far away from the truth. The conclusion is that communication, individual rationality and social deliberation are not absolute goods. Sometimes, especially where the aim is the truth, it is better that some agents effectively isolate themselves and avoid communication for some time.

3.5.2 Summary

This chapter has provided two modelings of the same example of collective failure: a "bad" informational cascade, leading everybody in a group to make the wrong choice between two alternatives:

- We have studied a case of informational cascade in two formal settings, corresponding to two different assumptions for the agents' rationality:
  - First, we have generalized the standard Bayesian analysis of cascades to show that such bad cascades also occur when rationality includes unbounded higher-order reasoning power, which is usually left out of the picture.
  - Second, we have shown that the fact that such bad cascades frequently occur in real societies cannot be imputed to the complexity of probabilistic reasoning. Indeed, we have proven that such cascades would

also occur if agents were simply assumed to count the number of pieces of evidence for each alternative and compare those two numbers.

- By doing this, we have given more ground to the somewhat counterintuitive claim that informational cascades are one of the collective consequences of individual rationality, even when they lead inescapably to a failure in society.

- After ruling out a flaw in the individuals' rationality, we have pointed towards another culprit: the intermediate situation in which agents find themselves: by sharing their decisions but not their justifications. Agents receive too much information not to be influenced by each other's choices, but not enough information to get the correct interpretation of those choices[10].

3.5.3 Further Research

Since a flaw in any of the individuals' reasoning is *not* a necessary condition for reaching collective failure, a natural aim for further research is to understand what are the sufficient and necessary conditions for such failures, beyond the specific phenomenon of informational cascades. This would help reaching another more activist aim: to understand how to prevent or reverse such situations towards better collective outcomes.

The next chapter takes a step in this direction. We will turn to a case-study of a second social phenomenon exemplifying some level of collective failure, pluralistic ignorance. This will have two important advantages: first, it will allow us to see what those two phenomena have in common which might explain collective failure. Moreover, by discussing how to reach a "better" collective situation, we will have taken into account a parameter of social phenomena which we have left implicit so far but which will be at the center of the chapters to come: the social network structure.

[10]In this chapter we assumed that agents who communicate their guess are in fact truthful sources of information. This assumption can be weakened by moving away from the public announcement operators used in this chapter to soft information upgrades. This direction has been explored in [162].

Chapter 4

Pluralistic Ignorance

4.1 Introduction

The previous chapter modeled how a population of perfectly rational individuals may form wrong beliefs by getting caught in a rational herding process. This chapter continues our investigation into the dynamics of collective failures. It focuses on another surprising but well-documented social phenomenon, often described as a collective “ social comparison error” [92, p. 126]: *pluralistic ignorance*, a situation where all individuals of a group hold the same beliefs and display the same behavior, but all believe that the others’ beliefs differ from theirs.

By giving a minimal modeling of the phenomenon, we will be able to capture its irony: it is precisely because no agent is different from the others that they can persist in believing that they are. In other words, it is the uniformity of the population which lets its individuals believe in their singular difference. We will also give a characterization of the situations minimizing the change needed to dissolve this collective error into a state where no error is left.

Beyond specific examples of collective failures, our objective is to design a logical framework to capture real-life social influence phenomena and their corresponding opinion dynamics. To do so, we will build on the setting which Liu, Seligman, and Girard designed in [120], designed to model opinion change under peer pressure in social networks, introduced in Section 2.2.2. We will enrich this framework by introducing a distinction between the purely private sphere of agents, namely their mental states, and the public sphere of their observable behavior, i.e., what they seem to believe, or what beliefs they publicly express.[1] We will show how the resulting framework allows for modeling complex social phenomena, and in particular cases where agents are wrong about each other’s opinions, such as the paradigmatic case of pluralistic ignorance.

[1]This modeling of “two-layered agents” is also the first step towards the more general “n-layered” framework which we will present in Chapter 5.

4.1.1 Outline

We start by introducing the phenomenon of pluralistic ignorance (4.1.2) and recalling its peculiar dynamic properties (4.1.3). Section 4.2 extends the setting from [120] to allow distinguishing between *private* opinions and *expressed* opinions of agents: we explain why such a distinction is useful in a framework for social influence phenomena (4.2.1), we make some assumptions about how the dynamics of social influence affects our agents' public behavior (4.2.2), and we present our logical framework (4.2.3). Section 4.3 then shows how this newly defined "two-layered" social influence setting allows us to model the case of pluralistic ignorance and to account for its dynamic properties: its "stability" (4.3.1), and its "fragility" in (4.3.2). Finally, Section 4.4 summarizes the chapter, gives directions for further research (4.4.3), and concludes by remarking that the two types of collective failures considered so far – informational cascades and pluralistic ignorance – rely on a common trigger: agents receive either too much or not enough information about each other (4.4.1).

4.1.2 The Phenomenon of Pluralistic Ignorance

The term "pluralistic ignorance" originates in the social and behavioral sciences in the work Allport and Katz [128]. It can be roughly defined as a situation where each individual of a group believes that her private attitude towards a proposition or norm differs from the rest of the group members', even though everyone in the group acts identically. For instance, after a difficult lecture which none of the students understood, it can happen that none of them asks any question even though the teacher explicitly requested them to do so in case they did not understand the material. There are numerous examples of pluralistic ignorance in the social and psychological literature such as, in addition to this classroom example, drinking among college students, attitudes towards racial segregation, and many more.[2]

Even though different definitions have been given in the literature [115, 92, 125, 128, 52], we follow [52] and define pluralistic ignorance as *a collective discrepancy between the agents' private attitudes and their public behavior*, namely a situation where all the individuals of a group have the same private attitude towards an object of opinion φ (say a belief in φ), but publicly "display" a conflicting attitude towards φ (say a belief in $\neg\varphi$).

[2]An extensive study of the classroom phenomenon was done by Miller and McFarland [124]. In a study of college students, Prentice and Miller [134] found that most students believed that the average student was much more comfortable with alcohol norms than they themselves were. Fields and Schuman [77] conducted a similar study, which showed that on issues of racial and civil liberties most people perceived others to be more conservative than they actually were. O'Gorman and Garry [129] found a similar tendency among whites to overestimate other whites' support for racial segregation.

4.1.3 A *Stable* State of Collective Error?

From a dynamic perspective, pluralistic ignorance is often reported as being both a *robust* and *fragile* phenomenon. It is robust in the sense that, if nothing changes in the environment, the phenomenon might persist over a long period of time – the college students might keep obeying an unwanted drinking norm for generations. On the other hand, it is fragile in the sense that if just one agent announces her private belief, it may be enough to dissolve the phenomenon – if just one student of the classroom example starts to ask questions about the difficult lecture the rest of the students might soon follow. The two-layer definition of social influence which we develop below will allow us to explain how pluralistic ignorance may dissolve in a community by cascading effects and thus allow us to illustrate both its robustness and its fragility.

4.2 Modeling Opaque Agents

In [120], Liu, Girard and Seligman design a hybrid logic to model opinion change induced by social influence in a community (for more detail, see Section 2.2.2). This chapter builds on their framework, both formally and conceptually. Let us briefly recall its most relevant features. On the static level: at any given moment, each agent in the network is assumed to be in one (and only one) of three possible opinion states (relatively to an implicit object of opinion φ): either she believes it ($B\varphi$), or she believes that $\neg\varphi$ ($B\neg\varphi$), or she has no opinion about it: ($U\varphi$). On the dynamic level: opinion change under social influence is defined in terms of two possible changes. First, an agent *adopts* an opinion as an effect of *Strong Influence*, i.e. when her friends unanimously hold that opinion. And, second, an agent *drops* an opinion under Weak Influence, i.e when none of her friends hold her current opinion and some hold the opposite opinion. In all other cases, the agent's opinion does not change.

4.2.1 Objections to Transparency

As mentioned in Section 2.2.2, an advantage of the simplicity of the framework of [120] is that it makes it unproblematic to identify the stability and stabilization conditions of social-doxastic configurations, both of which can be characterized directly in the language of friendship and belief. However, this simplicity is pricey: even though this is not explicitly mentioned as such, it relies on a strong assumption: *agents' belief states are influenced directly by their friends' belief states.* Thus, either all agents have direct access to their friends' beliefs (as mind-readers would), or their observed behavior always reflects their private beliefs, i.e., there is no difference between what they *seem to believe* and what they *actually believe.* This *transparency* assumption (all agents always automatically know what their friends/neighbors believe) prevents capturing any social situations where agents act in a way which does *not* reflect their mental states.[3]

Similar objections to the transparency assumption arise when discussing other mental states. Consider for instance preferences instead of beliefs. Let us assume that preferences are subject to social influence in a similar way, as in [166], in the sense that if all of my friends prefer option A to option B, I end up favoring A too. Intuitively, it seems that if I end up wearing a hat rather than none, chances are that it is not directly because all of my friends privately prefer to wear one, but rather because they *act as if* they do. For all I know, they could all be pretending because they all observe that everybody else is wearing a hat, and everyone could be following a trend that nobody actually likes.

Our main point here is that this distinction between private mental state and public behavior seems to be a key component of social sciences. Consider the documented cases where agents have been collectively enforcing a norm, for instance a segregation norm, despite the fact that they individually do not agree with it. It is precisely because they do *not* have access to each other's preferences and opinions that a collective behavior can result which goes against the ones of most (or all) agents, considered individually.[4] Therefore, it seems reasonable to require that a framework which aims at modeling social influence accommodates "opaque" agents.

Simply put, the situation of *pluralistic ignorance* on which this chapter focuses can be considered as an extreme case of "anti-transparency", since it involves all agents being wrong about each other's beliefs. Below, we will adapt the notion of opinion change under social influence from [120] to propose a "two-layer" notion to represent how "opaque" agents influence each other's behavior. We will then check whether this minimal two-layer adaptation captures the dynamic properties usually assigned to pluralistic ignorance: stability and fragility.

4.2.2 Opaque Social Influence

To reflect the fact that agents do *not* have access to what others privately believe, we introduce a distinction between *private belief*, which we name "inner belief" (I_B) and *public (or observable) behavior*, which we name "expressed belief" (E_B) Following [120], we define *two* notion of undecidedness or "unbelief" in the

[3]Such an additional "layer" of agents representation is also necessary for cases involving higher-order beliefs, since the complexity of such cases usually arises precisely from the fact that there might be a difference between what agent a believes that agent b believes and what agent b actually believes. However, unlike in Chapters 3 and 7, we will not pursue the issue of higher-order beliefs any further in this chapter. The two layers setting that we will use in this chapter could be entirely interpreted as distinguishing what an agent privately believes and what the others privately believe that she privately believes. Even though we might be interested of designing a similar framework using this distinction (*my* first order versus *their* higher-order belief), here we will distinguish merely between private belief and observable behavior.

[4]See for instance [144] on this issue.

obvious way:

$$I_U\varphi := \neg I_B\varphi \land \neg I_B\neg\varphi \qquad (\textit{inner} \text{ unbelief})$$

$$E_U\varphi := \neg E_B\varphi \land \neg E_B\neg\varphi \qquad (\textit{expressed} \text{ unbelief})$$

To define our new social influence operator, we will make the following simplifying assumption: from the subjective perspective of each agent, what matters (what influences her) is *what she herself privately believes and what the others seem to believe*. This reflects the fact that influence occurs (at least in good part) at the behavioral (observable, visible, displayed) level. We therefore adapt the notions of strong and weak influence from [120] (described in Section 2.2.2) accordingly:

> *2-layer strong influence* ($SI^2\varphi$) is the situation where all (and some) of my friends express the belief that φ: $SI^2 := FE_B\varphi \land \langle F\rangle E_B\varphi$.
>
> *2-layer weak influence* ($WI^2\varphi$) is the situation where some of my friends express the belief that φ and none of them expresses the belief that $\neg\varphi$: $WI^2 := \langle F\rangle E_B\varphi \land F\neg E_B\neg\varphi$.

We want to define social influence dynamics according to which my (expressed) reaction depends on asymmetric information: what *I privately believe* and what *the others express*. This reflects the fundamental asymmetry between the first and third person perspectives which is needed to model pluralistic ignorance. It is symmetric in that everybody reacts in the same way and in that everybody interprets the behavior of others in the same way; but it is asymmetric in that people don't have access to others' mental states and have a "privileged" access to their own.

Figure 4.1 lists the 24 possible situations of an individual among her friends, from her perspective, and describes her reaction (on the expressed level).[5] Her private attitude appears in the first column, the possible repartition of her friends' behaviors (expressed belief states) in columns 2,3,4 (in a truth table format – 1 for "true" and 0 for "false"), and her resulting behavior in one of the last three columns (depending of which type of agents we are considering). It is easy to see that our strong influence (rows 10 to 12 and 16 to 18 of the table) is still similar to the one from [120] but defined on the level of "expressed belief" instead of what was simply called "belief". However, weak influence (when not strong, rows 7 to

[5]The notation used in this chapter reflects the one inherited from the setting of [120], as introduced in Section 2.2.2. However, for simplicity, the next chapters we will adopt the more concise notation used in our later publications: "$I_B\varphi$" will be replaced by "*ip*" (standing for "inner pro opinion");"$I_B\neg\varphi$" will be replaced by "*ic*" ("inner contra opinion"); $I_U\varphi$ will be replaced by "*in*" ("inner neutral opinion"),"$E_B\varphi$" will be replaced by "*ep*" ("expressed pro opinion"), "$E_B\neg\varphi$" will be replaced by "*ec*" ("expressed contra opinion"), and $E_U\varphi$ will be replaced by "*en*" ("expressed neutral opinion").

	Inner state	$\langle F\rangle E_B\varphi$	$\langle F\rangle E_B\neg\varphi$	$\langle F\rangle E_U\varphi$	Type 1	Type 2	Type 3
1	$I_B\varphi$				$\rightsquigarrow E_B\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_B\varphi$
2	$I_B\neg\varphi$	1	1	1	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_B\neg\varphi$
3	$I_U\varphi$				$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$
4	$I_B\varphi$				$\rightsquigarrow E_B\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_B\varphi$
5	$I_B\neg\varphi$	1	1	0	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_B\neg\varphi$
6	$I_U\varphi$				$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$
7	$I_B\varphi$				$\rightsquigarrow E_B\varphi$	$\rightsquigarrow E_B\varphi$	$\rightsquigarrow E_B\varphi$
8	$I_B\neg\varphi$	1	0	1	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$
9	$I_U\varphi$				$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$
10	$I_B\varphi$						
11	$I_B\neg\varphi$	1	0	0	$\rightsquigarrow E_B\varphi$	$\rightsquigarrow E_B\varphi$	$\rightsquigarrow E_B\varphi$
12	$I_U\varphi$						
13	$I_B\varphi$				$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$
14	$I_B\neg\varphi$	0	1	1	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_B\neg\varphi$
15	$I_U\varphi$				$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_B\neg\varphi$
16	$I_B\varphi$						
17	$I_B\neg\varphi$	0	1	0	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_B\neg\varphi$
18	$I_U\varphi$						
19	$I_B\varphi$				$\rightsquigarrow E_B\varphi$	$\rightsquigarrow E_B\varphi$	$\rightsquigarrow E_U\varphi$
20	$I_B\neg\varphi$	0	0	1	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_U\varphi$
21	$I_U\varphi$				$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$
22	$I_B\varphi$				$\rightsquigarrow E_B\varphi$	$\rightsquigarrow E_B\varphi$	$\rightsquigarrow E_B\varphi$
23	$I_B\neg\varphi$	0	0	0	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_B\neg\varphi$	$\rightsquigarrow E_B\neg\varphi$
24	$I_U\varphi$				$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$	$\rightsquigarrow E_U\varphi$

Figure 4.1: Influence on three different types of agents.

9 and 15 to 18) now results in a different state depending on the initial *private* belief state of the agent herself (see for instance rows 7 and 8).

There are two possible cases in which I have friends (unlike in rows 22 to 24) but I am neither strongly nor weakly influenced: whenever all of my friends express undecidedness (rows 19 to 21) and whenever some of them express the belief that φ while some express the belief that $\neg\varphi$ (rows 1 to 6). In the setting of [120], nothing happens, i.e, the agent continues to believe whatever she did before. In our setting, we have to make a choice as to what the agent expresses. The simplest one is to assume that in both these cases, agents express their true private belief (act sincerely). This corresponds to agent of type 1 in Figure 4.1, which is entirely determined by the following instructions:

4.2.1. Definition. [2-layer social influence (agents of type 1)] The following rules constitute the notion of *2-layer social influence*:

- $SI^2\varphi$: If an agent is in a situation of strong influence with φ, she will express the belief that φ ($E_B\varphi$) at the next moment, whatever her current (inner and expressed) state.

- $WI^2\varphi$: If an agent is in a situation of weak influence with φ she will express the belief that φ ($E_B\varphi$) at the next moment if she is currently privately undecided about φ ($I_U\varphi$) or if she already privately believes that φ ($I_B\varphi$), and she will express undecidedness ($E_U\varphi$) if she privately believes $\neg\varphi$ ($I_B\varphi$).

- Otherwise, the agent will express at the next moment her private current state.

However, some agents might be more inclined to follow the others, and they do so in different ways. Types 2 and 3 in the table are examples of other possible types of agents which still comply with our definition of two-layer strong and weak influence. If I am a type 2 agent, I will be sincere (i.e., my expressed belief state will correspond to my inner belief state) whenever I face no opposition[6]. For instance, if I privately believe that φ, I will express this belief if none of my friends expresses a belief in $\neg\varphi$. And if I am a type 3 agent, I will be sincere whenever some of my friends express support for my private belief state, I will for instance express my inner belief in φ if some of my friends express a belief in φ too. Type 1 agents are thus simply the ones that are sincere in both cases: when they get some support and when they face no opposition.

We will see in section 4.3 how the dynamic properties of social phenomena like pluralistic ignorance depend on the type of agents involved but let us first introduce the tools we will use to represent changes of the (two-layered) state of agents in a social network.

4.2.3 Two-layer Opinion Change Logic

In this section, we briefly introduce an extension of the setting of [120] to reason about the influence dynamics for "two-layered" agents, which will allow us to model cases such as pluralistic ignorance. We start with a static logic and then move on to give the full dynamics.[7]

In section 4.2 we introduced the two relevant properties of each agent which we want to model, namely her inner (private) belief state and her expressed belief state. We assume that each of the two properties always takes exactly

[6]Note that the notion of sincerity we have adopted matches the one used in [25, 135], where "an assertion of P is sincere if the speaker believes that P". Later this concept received a more generalized interpretation in [31]. In this thesis we adopt the first reading, hence an agent is sincere if her expressed belief state will correspond to her inner belief state.

[7]The next chapter will present a generalization of the setting to be defined here with its Hilbert-style proof system and completeness result. We therefore leave out the very similar discussion of soundness and completeness for the particular case presented here. What we want to focus on in this chapter is the conceptual necessity for any framework modeling social influence to account for the distinction between (at least) the private mental states of agents and their observable behavior, and to show how this simple distinction allows capturing the dynamics of well-documented social phenomena, by focusing on the case of pluralistic ignorance.

one of the three same values proposed by [120]: belief, disbelief, or undecidedness. Our language will therefore contain six atomic corresponding propositions: $\mathsf{PROP} = \{I_B\phi, I_B\neg\phi, I_U\phi, E_B\phi, E_B\neg\phi, E_U\phi, \}$. We assume a countable infinite set of nominals (NOM) used as names for agents in possible networks (just as they are used to name states in traditional hybrid logic [10]). The syntax of our static language is given by:

$$\varphi \ ::= \ p \mid i \mid \neg\varphi \mid \varphi \wedge \varphi \mid F\varphi \mid G\varphi \mid @_i\varphi \ ,$$

where $p \in \mathsf{PROP}$ and $i \in \mathsf{NOM}$.[8]

Models. A model is a distribution of private and expressed opinions to each agent in a symmetric and connected network. More precisely, a model is a tuple $\mathcal{M} = (A, \sim, g, \nu)$, where A is a non-empty set of agents, $\sim$ is a binary relation on A representing the network structure[9], $g : \mathsf{NOM} \to A$ is a function assigning an agent to each nominal, and $\nu : A \to \mathsf{PROP}$ is a valuation, with the particular constraint that it assigns exactly two elements of PROP to each $a \in A$: one element of $\{I_B\phi, I_B\neg\phi, I_U\phi\}$ and one element of $\{E_B\phi, E_B\neg\phi, E_U\phi\}$.

To represent the change of opinions due to social influence, we add to our language a dynamic modality $[\mathcal{I}]$ in the way of standard Dynamic Epistemic Logic [24, 70]. Given a model $\mathcal{M} = (A, \sim, g, \nu)$ and an influence event $\mathcal{I}$, the updated model is given by $\mathcal{M} \otimes \mathcal{I} = (A, \sim, g, \nu')$, where ν' assigns to each $a \in A$: the same element of $\{I_B\phi, I_B\neg\phi, I_U\phi\}$ as ν and the element $\{E_B\phi, E_B\neg\phi, E_U\phi\}$ is the one given in the table in Figure 4.1 (for a given agent type among the three types proposed).[10]

Given a $\mathcal{M} = (A, \sim, g, \nu)$, an $a \in A$, $p \in \mathsf{PROP}$, an operator $\mathcal{I}$ defined by the table in Figure 4.1 and a formula φ of our language, we define the truth of φ at

[8] We will use the standard abbreviations for disjunction, material implication, and equivalence ($\vee$, $\to$, and $\leftrightarrow$) and denote the dual operator of F by $\langle F\rangle$ and the dual of G by $\langle G\rangle$. The intuitive meaning of the F and $@_i$ operators were already discussed in Section 2.2.2. The G-operator is the global modality quantifying over all agents in the network and $G\varphi$ is read as "all agents (satisfy) φ".

[9] If we are talking about undirected networks, we will assume that $\sim$ is symmetric.

[10] There are more possible influence operators to be defined in terms of combination of inner and expressed opinions. However, this chapter restricts itself to illustrating how some given notion of social influence can be extended to model the dynamic properties of pluralistic ignorance. The next chapter will generalize to agents with more than two properties and to all definable ways of updating the corresponding models.

a in $\mathcal{M}$ inductively by:

$$
\begin{array}{lll}
\mathcal{M}, a \models p & \text{iff} & p \in \nu(a) \\
\mathcal{M}, a \models i & \text{iff} & g(i) = a \\
\mathcal{M}, a \models \neg\varphi & \text{iff} & \text{it is not the case that } \mathcal{M}, a \models \varphi \\
\mathcal{M}, a \models \varphi \wedge \psi & \text{iff} & \mathcal{M}, a \models \varphi \text{ and } \mathcal{M}, a \models \psi \\
\mathcal{M}, a \models G\varphi & \text{iff} & \text{for all } b \in A; \mathcal{M}, b \models \varphi \\
\mathcal{M}, a \models F\varphi & \text{iff} & \text{for all } b \in A; a \sim b \text{ implies } \mathcal{M}, b \models \varphi \\
\mathcal{M}, a \models @_i\varphi & \text{iff} & \mathcal{M}, g(i) \models \varphi \\
\mathcal{M}, a \models [\mathcal{I}]\varphi & \text{iff} & \mathcal{M} \otimes \mathcal{I}, a \models \varphi
\end{array}
$$

Satisfiability, validity etc. are defined as usual.

4.3 Pluralistic Ignorance Revisited

We will use the idea developed in the previous sections to model pluralistic ignorance. We assume that everyone in a "group" is connected to everyone else through some finite number of steps (a "community" in the sense of [145]), and that the social relation. In other words, we will work with connected network models. Moreover, we will assume that the $\sim$ relation is symmetric in the rest of the section.

Pluralistic ignorance, in the sense that everybody inner believes φ but expresses a belief in $\neg\varphi$, can be formalized by:

$$PI\varphi := G(I_B\varphi \wedge E_B\neg\varphi) \tag{4.1}$$

If $PI\varphi$ is true in a network model $\mathcal{M}$ we will say that $\mathcal{M}$ *is in a state of pluralistic ignorance*.

To investigate how social influence affects pluralistic ignorance we need to define the event model that captures the two-layer influence described in Section 4.2. This is fairly straightforward given the table of Figure 4.1. For now we assume that all agents are of type 1 mentioned in Section 4.2. We will return to considering other types of agents later on. For each of the 24 rows, the conjunction of the first four columns will be a precondition. These 24 preconditions will clearly be pairwise inconsistent. For instance, the fourth row gives the precondition formula

$$I_B\varphi \wedge \langle F\rangle E_B\varphi \wedge \langle F\rangle E_B\neg\varphi \wedge \neg\langle F\rangle E_U\varphi.$$

The corresponding post-condition will be the assignment assigning $B\varphi$ to the inner state of the agent and $B\varphi$ to her express state, as specified by the first and the fifth column of the table. The resulting event model will be denoted $\mathcal{I}$.

4.3.1 Stability

As claimed in Section 4.2, pluralistic ignorance constitutes a "robust" state, or "equilibrium", in the sense that if a network is in a state of pluralistic ignorance it will stay in this state. This turns out to be trivially true (and this for all three types of agents defined). We formalize this in the following lemma:

4.3.1. Proposition. *A connected network model in a state of pluralistic ignorance is stable and the condition for being stable reduces to*

$$PI\varphi \to [\mathcal{I}]PI\varphi. \tag{4.2}$$

Proof: If a network model $\mathcal{M}$ satisfies $PI\varphi$, then by definition every agent satisfies $I_B\varphi$ and $E_B\neg\varphi$. A simple inspection of row 16 in the table of Figure 4.1 shows that all agents will keep expressing a belief in $\neg\varphi$ and keep their inner belief in φ after an update with $\mathcal{I}$. Thus, $PI\varphi$ will remain true after the update, i.e. $[\mathcal{I}]PI\varphi$ is true, hence the situation is stable. □

4.3.2 Fragility

The "fragility" component of pluralistic ignorance is a little more complex. If just one agent announces her private belief this may "dissolve" the phenomenon into a radically opposite state of collective sincerity, or not, depending on the structure of the network, as we will show. More precisely, we take pluralistic ignorance (in the form of (4.1)) to be *dissolved* when it is true that $G(I_B\varphi \wedge E_B\varphi)$. Assume that the network model $\mathcal{M}$ is in a state of pluralistic ignorance, i.e. $\mathcal{M}$ satisfies $PI\varphi$. Now assume that some agent (maybe by mistake) suddenly expresses her true inner belief in φ. Let us refer to this agent by the nominal i. Then the following is now satisfied in $\mathcal{M}$

$$UPI\varphi := @_i(I_B\varphi \wedge E_B\varphi) \wedge G\big(\neg i \to (I_B\varphi \wedge E_B\neg\varphi)\big).$$

A model satisfying $UPI\varphi$ (where i might be replaced by another nominal) will be said to be in a state of *unstable pluralistic ignorance.*[11] How $\mathcal{M}$ will evolve under iterated applications of the influence event $\mathcal{I}$ depends on several factors.

First, consider the case where i will keep expressing her true belief.[12] Then, if $\mathcal{M}$ is connected (and finite) it is easy to show that after a finite number of updates by the influence event $\mathcal{I}$, $\mathcal{M}$ will end up in a stable state where everyone

[11] Note that the name of "unstable pluralistic ignorance" used here does not refer to a particular case of pluralistic ignorance, but to a state of "almost" pluralistic ignorance, a state which minimally differs from it at the observable level, by one agent expressing her private beliefs sincerely, when nobody else does.

[12] Formally, remark that we have to make a small change to $\mathcal{I}$ to make sure that i will not change her expressed belief after one step.

expresses their true beliefs: By inspecting row 4 in the table of Figure 4.1, it follows that after one update by $\mathcal{I}$ all of i's friends will express a belief in φ and that after another update with $\mathcal{I}$ the friends of friends of i will also express their true belief. In this way, a cascade effect will spread the change throughout the network and result in a stable state where everyone expresses the same true belief.

However, this was assuming the involvement of a special agent who would not obey the rules of social influence we have assumed others to obey. If i is only assumed to express her true belief for a *single* round, and then be subject to social influence as anybody else, things get more complicated. She will, in the next round already, revert to expressing a belief in $\neg\varphi$ by the effect of social influence (as all of i's friends originally expressed a belief in $\neg\varphi$). For this reason, other agents might be influenced back and forth too, and the distribution might keep "fluctuating" and never stabilize. Here is an example of the later case, where i refers to agent a[13]:

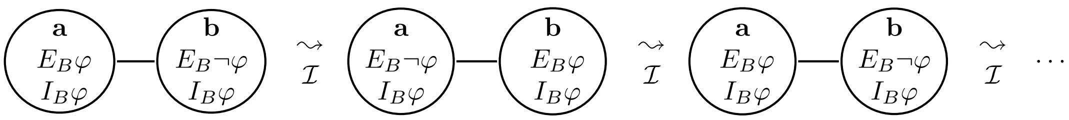


Figure 4.2: Two agents a and b with opposite expressed opinions influencing each other, resulting in their expressed opinions oscillating.

The above example shows that a state of unstable pluralistic ignorance will not necessarily stabilize, and hence not necessarily result in a state where pluralistic ignorance is dissolved. Below, we give a characterization of the ones which do result in such a state, given our assumption that all agents are of type 1.

4.3.2. Proposition. *Let $\mathcal{M} = (A, \sim, g, \nu)$ be a finite, connected, symmetric network model in a state of unstable pluralistic ignorance. Then the following are equivalent:*

(i) After a finite number of updates by the influence event $\mathcal{I}$, $\mathcal{M}$ will end up in a stable state where pluralistic ignorance is dissolved, i.e. there is a $k \in \mathbb{N}$ such that $\mathcal{M} \otimes^k \mathcal{I} \models G(I_B\varphi \wedge E_B\varphi)$ and $\mathcal{M} \otimes^k \mathcal{I} = \mathcal{M} \otimes^{k+1} \mathcal{I}$.

(ii) There is an agent that expresses her true belief in φ for two rounds in a row, i.e. there is an $a \in A$ and a $k \in \mathbb{N}$ such that $\mathcal{M} \otimes^k \mathcal{I}, a \models E_B\varphi$ and $\mathcal{M} \otimes^{k+1} \mathcal{I}, a \models E_B\varphi$.

(iii) There are two agents that are friends and both express their true beliefs in φ in the same round, i.e. there are $a, b \in A$ and a $k \in \mathbb{N}$ such that $a \sim b$, $\mathcal{M} \otimes^k \mathcal{I}, a \models E_B\varphi$, and $\mathcal{M} \otimes^k \mathcal{I}, b \models E_B\varphi$.

[13]Here we regain the same fluctuation case that was given in [120], except that it now occurs, as wanted, at the level of expressed belief instead of "belief".

(iv) There are two agents that are friends and have paths of the same length to the agent named by i, i.e. there are agents $a, b \in A$ and a $k \in \mathbb{N}$ such that $a \sim b$, $\mathcal{M}, a \models \langle F \rangle^k i$, and $\mathcal{M}, b \models \langle F \rangle^k i$.

(v) There is a cycle in $\mathcal{M}$ of odd length starting at the agent named by i, i.e. there is a $k \in \mathbb{N}$ such that $\mathcal{M} \models @_i \langle F \rangle^{2k-1} i$.

(vi) There is a cycle in $\mathcal{M}$ of odd length, i.e. there is a $k \in \mathbb{N}$ and $a_1, a_2, ..., a_{2k-1} \in A$ such that $a_1 \sim a_2, a_2 \sim a_3, ..., a_{2k-2} \sim a_{2k-1}, a_{2k-1} \sim a_1$.

A proof of this proposition can be found in Appendix (4.5).

By this proposition (and its proof) we can also come up with an upper bound of the number of update-steps needed for a network model in an unstable pluralistic ignorance state to dissolve, if it stabilizes. If a network model $\mathcal{M} = (A, \sim, g, \nu)$ stabilizes it follows from (iv) that there are $a, b \in A$ and a $k \in \mathbb{N}$ such that $a \sim b$ and a and b both have a path of length k to $g(i)$. Choose the smallest such k. For all $c \in A$, let $m(c)$ be the length of the shortest path to either a or b. Then, by inspecting the proof it is not hard to see that $\mathcal{M}$ stabilizes in a state where pluralistic ignorance is dissolved in at most $k + max_{c \in A}\{m(c)\}$ steps.

4.3.3 Comparing Types of Agents

As mentioned in section 4.2, the type of agents might also influence whether unstable pluralistic ignorance will dissolve. In the above we have focused on what happens when agents are of type 1. If one wants all agents to be of another type, then one can simply change the definition of $\mathcal{I}$. First, note that agents in a state of pluralistic ignorance will always be strongly influenced and since all the three different kinds of agents react the same to strong influence, Proposition 4.3.1 remains true for all types.

Now, let us consider a network of type 3 agents (expressing their inner belief whenever they have some support for it). The lines 1, 4, 7, and 10 of Figure 4.1 will stay unchanged. Thus, Proposition 4.3.2 will remain true for this type of agents. The only case left to consider is therefore whether Proposition 4.3.2 holds for type 2 agents (expressing their inner belief whenever they face no opposition). We leave this as an open problem.

4.4 Conclusion

4.4.1 Diagnosis: Too Much (or Not Enough) Information!

By modeling a well-known phenomenon of collective inefficient behavior, and by showing that this modeling allows capturing its dynamic properties, we have illustrated why the distinction between "visible" and "invisible" properties of agents is essential to their social condition, and hence to the dynamics of social phenomena. If agents had access to each other's private opinions, they would know that they have some support from their network-neighbors and pluralistic ignorance could simply not occur. Hence, the dynamics of social influence effects seems to rely on the intermediate state agents find themselves in: agents are influenced only by what they can observe from each other.

Similarly as in the case of informational cascades discussed in the previous chapter, the very possibility of a collective error such as pluralistic ignorance relies on the condition that people have access to *some* information about each other, but not to *all* relevant information. In the case of informational cascades, agents have access to the choices of others but not to the reasons justifying these choices (and if they had, a cascade leading to the wrong choice could not occur). In the case of pluralistic ignorance, agents can observe each other's behavior but not each other's private opinion (and if they could, there would be no ignorance).

4.4.2 Summary

This chapter presented a "two-layer" framework for opinion change under social influence in networks. Let us summarize its main findings:

- We have argued that a distinction between what agents privately believe and what they display to their network-neighbors is essential to the dynamics of social influence phenomena, and in particular to phenomena of "collective failure", such as pluralistic ignorance.

- We have extended the framework from Liu, Seligman, and Girard [120] to a "two-layer" framework for opinion change over social networks.

- We have proposed a definition of social influence (for three types of agents) relying on the following assumptions: 1) agents influence each other not directly by what they privately believe but by what they publicly express; 2) what agents will express next depends on asymmetric information: what their neighbors express and what they themselves privately believe, and 3) agents tend to express their private opinions sincerely, unless social pressure prevents them too.

- We have shown how those assumptions were sufficient to explain the dynamic properties of pluralistic ignorance as they are usually described by social psychologists: its stability, as well as its fragility.

- We have given a characterization of the network configurations for which a unique agent expressing her opinion sincerely would induce a "sincerity cascade", resulting in everybody sincerely expressing their opinion.

- Finally, we have observed that, similarly as in the case of informational cascades, the collective failure involved in pluralistic ignorance is the result of agents receiving some information about each other but not enough information to prevent misinterpreting each other's behavior.

4.4.3 Further Research

Mixing types of agents. An interesting case for further research would be networks with *mixed* types of agents. Our framework can be used to model this as well. We simply add another variable to the modeling of agents (in addition to their inner and expressed belief states) to keep track of the agents' types, which we assume to always be of exactly one of the three types. We can then modify the definition of the social influence operator $\mathcal{I}$ such that in the lines where the agent's type affects what they will do we split each line into three new lines distinguished by the extra preconditions of the form "type 1" (or 2 or 3). Then we change the corresponding post-conditions accordingly. In this way, a new event model $\mathcal{I}'$ can be defined, resulting in an influence dynamics that also depends on the agents types. We will leave the details of this for future research.

Resolving inner conflicts. Even though we have shown that pluralistic ignorance is stable, psychology also suggests that the phenomenon will not continue forever. The discrepancy between one's inner beliefs and one's expressed beliefs is a conflict which might have negative consequences for the agents and as such they may very well try to resolve it. This is a well studied issue in the social and psychological literature on pluralistic ignorance. It is usually assumed [134] that the agents have three different ways in which they can act to resolve this conflict. They can either *internalize the perceived view of their peers*, i.e. change their private beliefs, *attempt to change the perceived view of their peers*, or *alienate themselves form their peers*. In our setting, the first option simply corresponds to the agents changing their inner beliefs in φ to inner beliefs in $\neg\varphi$. the only way they can try and change the opinion of others is by their expressed belief. Thus, the most natural interpretation of the second option would be that the agents will start expressing their true beliefs in φ. Finally, one interpretation of the action of alienating oneself from one's peers would be to remove friendship links to all agents that express a belief in $\neg\varphi$.

Different agents might choose different reactions to a conflict between their inner and expressed beliefs. Therefore, it would be natural to add an additional layer to keep track of what action an agent will chose in case of such a conflict. Moreover, it would be natural to assume that agents only try to eliminate this conflict after experiencing it for some time, i.e., for a given number of rounds. We could also capture this by adding another variable that acts as a "counter" of rounds. These new variables can then be included in the preconditions of the influence event $\mathcal{I}$. For the first two options it is obvious what the new post-conditions should be, but for the third option we need an extension of our notion of event model such that it can also change the links in a network model. We believe this can be done, but we leave the details for future research.

Conclusion to Part II

By modeling two well-known social phenomena of collective failure, this part of the thesis has given some insight into what a logic for social influence should take into account. The distinction between "visible" and "invisible" properties of agents is essential to their social condition, and hence to the dynamics of social phenomena over networks. As we have shown, both cases of failure rely on the condition that people have access to *some* information about each other, but not to *all* relevant information. Therefore, a logical framework for social influence dynamics should be able to capture the agents' "sight", in the sense of what they know about each other. We have provided a logical framework which can do just that.

In the next part of the thesis, we will continue our quest for a logical framework for social influence dynamics. First, generalizing beyond the "two-layer" picture of agents and beyond the case of *opinion* diffusion, we will introduce a "many-layer" logical framework to capture a much wider range of diffusion phenomena in social networks. And second, we will formally capture what is "visible" and "invisible" to agents in a social network by bringing full-fledged epistemic logic back into the picture. This will allow us to reason about the interaction between information and social influence dynamics in a more general perspective.

4.5 Appendix: Proof of Proposition 4.3.2

We give the proof of Proposition 4.3.2. The following lemma is a first step:

4.5.1. Lemma. *The following is a validity of our logic for all $j \in \mathsf{NOM}$ and all $k \in \mathbb{N}$:*

$$\big(@_i(I_B\varphi \wedge E_B\varphi) \wedge GI_B\varphi \wedge @_i\langle F\rangle^k j\big) \to [\mathcal{I}]^k @_j E_B\varphi. \tag{4.3}$$

Intuitively, this lemma says that if i for one round expresses her private belief in φ, if everyone else privately believes φ as well, and if there is a path from i to j of length n, then after exactly n updates with $\mathcal{I}$, j will express a belief in φ.

Proof: The proof goes by induction on $k \in \mathbb{N}$. By inspecting the lines 1, 4, 7 and 10 of table of Figure 4.1 the validity for $k = 1$ easily follows. Now assume that (4.3) is true for a $k \in \mathbb{N}$ and all $j \in \mathsf{NOM}$. Assume furthermore that the antecedent is true for $k+1$ in a model $\mathcal{M} = (A, \sim, g, \nu)$. This means that there is a path of length $k+1$ from $g(i)$ to $g(j)$, in particular there is an agent a such that there is path of length k from $g(i)$ to a and a path of length 1 from a to $g(j)$. We can assume without loss of generality that there is a nominal $h \in \mathsf{NOM}$, different from i and j, such that $g(h) = a$. But then by the assumption that (4.3) is true for k we obtain that $[\mathcal{I}]^k @_h E_B\varphi$ is true. But then by inspecting the lines 1, 4, 7, and 10 of table of Figure 4.1 it follows that $[\mathcal{I}]^{k+1} @_j E_B\varphi$ is true, as well. This completes the proof. $\Box$

The following lemma will also be useful:

4.5.2. Lemma. *Let $\mathcal{M} = (A, \sim, g, \nu)$ be a finite, connected, symmetric network model in a state of unstable pluralistic ignorance. Then for all $k \in \mathbb{N}_0$,*

$$\mathcal{M} \otimes^k \mathcal{I} \models G(E_B\varphi \vee E_B\neg\varphi).$$

Thus, when a network starts out in an unstable state of pluralistic ignorance and evolves under the influence event $\mathcal{I}$, no one will ever express undecidedness.

Proof: Assume that $\mathcal{M} = (A, \sim, g, \nu)$ is a finite, connected, symmetric network model in a state of unstable pluralistic ignorance. First note that since $\mathcal{M}$ is in a state of unstable pluralistic ignorance all agents satisfy $I_B\varphi$ and as the influence event $\mathcal{I}$ does not change any agent's inner belief, $I_B\varphi$ will remain true of all agents in all models of the form $\mathcal{M} \otimes^k \mathcal{I}$.

The proof goes on induction on $k \in \mathbb{N}_0$. The induction follows trivially from the fact that $\mathcal{M} \models @_i(I_B\varphi \wedge E_B\varphi) \wedge G\big(\neg i \to (I_B\varphi \wedge E_B\neg\varphi)\big)$. Now assume that $\mathcal{M} \otimes^k \mathcal{I} \models G(E_B\varphi \vee E_B\neg\varphi)$ and consider an agent $a \in A$. Note that all a's friends either expressed a belief in φ or a belief in $\neg\varphi$ in $\mathcal{M} \otimes^k \mathcal{I}$. But then, by inspecting the lines 4, 10, and 16 of the table of Figure 4.1, a must either express a belief in φ or a belief in $\neg\varphi$ in $\mathcal{M} \otimes^{k+1} \mathcal{I}$. This completes the induction proof. □

We can now prove the main proposition.

Proof of Proposition 4.3.2. $(i) \Rightarrow (ii)$. This is straightforward.

$(ii) \Rightarrow (iii)$. Assume that there are $a \in A$ and $k \in \mathbb{N}$ such that $\mathcal{M} \otimes^k \mathcal{I}, a \models E_B\varphi$ and $\mathcal{M} \otimes^{k+1} \mathcal{I}, a \models E_B\varphi$. By Lemma 4.5.2 and an inspection of the lines 4 and 10 of the table of Figure 4.1, it follows that there is a $b \in A$ such that $a \sim b$ and $\mathcal{M} \otimes^k \mathcal{I}, b \models E_B\varphi$. Hence, (iii) follows.

$(iii) \Rightarrow (iv)$. Let $a, b \in A$ and $k \in \mathbb{N}$ be such that $a \sim b$, $\mathcal{M} \otimes^k \mathcal{I}, a \models E_B\varphi$, and $\mathcal{M} \otimes^k \mathcal{I}, b \models E_B\varphi$. As previously, it follows by Lemma 4.5.2 and an inspection of the lines 4 and 10 of the table of Figure 4.1, that there are $a_{k-1}, b_{k-1} \in A$ such that $a_{k-1} \sim a$, $b_{k-1} \sim b$, $\mathcal{M} \otimes^{k-1} \mathcal{I}, a_{k-1} \models E_B\varphi$, and $\mathcal{M} \otimes^{k-1} \mathcal{I}, b_{k-1} \models E_B\varphi$. But then, it follows by Lemma 4.5.2 and an inspection of the lines 4 and 10 of the table of Figure 4.1, that there are $a_{k-2}, b_{k-2} \in A$ such that $a_{k-2} \sim a_{k-1}$, $b_{k-b} \sim b_{k-1}$, $\mathcal{M} \otimes^{k-2} \mathcal{I}, a_{k-2} \models E_B\varphi$, and $\mathcal{M} \otimes^{k-2} \mathcal{I}, b_{k-2} \models E_B\varphi$. Continuing this way, we obtain $a_0, a_1, a_2, ..., a_{k-1} \in A$ and $b_0, b_1, b_2, ..., b_{k-1} \in A$ such that $a_0 \sim a_1 \sim ... \sim a_{k-1} \sim a$, $b_0 \sim b_1 \sim ... \sim b_{k-1} \sim b$, $\mathcal{M} \otimes^0 \mathcal{I}, a_0 \models E_B\varphi$, and $\mathcal{M} \otimes^0 \mathcal{I}, b_0 \models E_B\varphi$. Now since $\mathcal{M} \otimes^0 \mathcal{I} = \mathcal{M}$, and $g(i)$ is the only agent in $\mathcal{M}$ that satisfy $E_B\varphi$, it follows that $a_0 = b_0 = g(i)$. Thus, $\mathcal{M}, a \models \langle F\rangle^k i$ and $\mathcal{M}, b \models \langle F\rangle^k i$ and (iv) follows.

$(iv) \Rightarrow (v)$. Assume that there are $a, b \in A$ and $k \in \mathbb{N}$ such that $a \sim b$, $\mathcal{M}, a \models \langle F\rangle^k i$, and $\mathcal{M}, b \models \langle F\rangle^k i$. From $\mathcal{M}, a \models \langle F\rangle^k i$, it follows that there are $a_1, a_2, ..., a_{k-1} \in A$ such that $g(i) \sim a_1 \sim a_2 \sim ... \sim a_{k-1} \sim a$. Similar there are $b_1, b_2, ..., b_{k-1} \in A$ such that $g(i) \sim b_1 \sim b_2 \sim ... \sim b_{k-1} \sim b$. But then

$$g(i) \sim a_1 \sim ... \sim a_{k-1} \sim a \sim b \sim b_{k-1} \sim ... \sim b_1 \sim g(i)$$

is a path of length $2(k+1)-1$. Hence, $\mathcal{M}, g(i) \models \langle F\rangle^{2(k+1)-1} i$ and (v) follows.

$(v) \Rightarrow (vi)$. This is trivial.

$(vi) \Rightarrow (iv)$. Assume that there is an odd cycle in $\mathcal{M}$. Since $\mathcal{M}$ is connected, there is a path from $g(i)$ to the cycle. But then it is not hard to find agents a and b of the cycle such that they both have same path length to $g(i)$ and are friends. Now, (iv) easily follows.

$(iv) \Rightarrow (i)$. Assume that there are $a, b \in A$ and a $k \in \mathbb{N}$ such that $a \sim b$, $\mathcal{M}, a \models \langle F\rangle^k i$, and $\mathcal{M}, b \models \langle F\rangle^k i$. Then by Lemma 4.5.1 it is not hard to see that after k updates with $\mathcal{I}$ both a and b will express beliefs in φ, i.e. $\mathcal{M} \otimes^k \mathcal{I}, a \models E_B\varphi$ and $\mathcal{M} \otimes^k \mathcal{I}, b \models E_B\varphi$. Then, since $a \sim b$, an inspection of the lines 1, 4, 7, and 10 of table of Figure 4.1 shows that a and b will keep expressing their belief in φ, i.e. $\mathcal{M} \otimes^{k+m} \mathcal{I}, a \models E_B\varphi$ and $\mathcal{M} \otimes^{k+m} \mathcal{I}, b \models E_B\varphi$ for all $m \in \mathbb{N}$. For the same reasons, for all $c \in A$ such that either $c \sim a$ or $c \sim b$, $\mathcal{M} \otimes^{k+1+m} \mathcal{I}, c \models E_B\varphi$ for all $m \in \mathbb{N}$. Similarly, for all $d \in A$ such that there is a $c \in A$ such that $d \sim c \sim a$ or $d \sim c \sim b$, we have that $\mathcal{M} \otimes^{k+2+m} \mathcal{I}, d \models E_B\varphi$ for all $m \in \mathbb{N}$. Generally for all $d \in A$ such that there are $c_1, ..., c_l \in A$ such that $d \sim c_1 \sim c_2 \sim ... \sim c_l$, $c_l = a$

or c_l, we have that $\mathcal{M} \otimes^{k+l+m} \mathcal{I}, d \models E_B\varphi$ for all $m \in \mathbb{N}$. Since $\mathcal{M}$ is finite and connected, there will be a stage l such that all agents in A have a path of length less than l to a or b and thus they all express belief in φ and will continue doing this. This completes the proof of Proposition 4.3.2.

Part III

Diffusion Phenomena

Introduction to Part III:
From Diffusion to Information and Back

Beyond the specific case studies presented in Part II, we are interested in a logical perspective on diffusion and information phenomena in social networks in general. This part of the thesis will therefore abstract away from specific examples of opinion spread, moving further away from social psychology and getting closer towards social network analysis.

Typically, the social network analysis perspective on the spread of information, viruses, trends, opinions, or behaviors in social networks (see Section 2.1.2 and [127, 105, 71]), sees them as locally and uniformly driven: whether an agent adopts a behavior, opinion, disease, product, trend, etc., depends on whether the agents linked to her within her social network have adopted it already.

Given that such processes depend on *local* properties of agents on a network graph, it seems natural to aim to develop a complete dynamic modal logic to reason about such "contagion via network-neighbors" phenomena in their generality. The ambition here is to capture the general logical laws of diffusion and information processes over networks, for a wide variety of dynamic processes. This is exactly what we will do in this part of the thesis.

Chapter 5 presents our first general logical framework for reasoning about diffusion processes within social networks, in which many different types of dynamics can be "plugged-in". Later on, Chapter 6 will include epistemic logic components into this framework in order to capture the interaction between information and diffusion phenomena, while Chapter 7 will propose a dynamic epistemic logic for diffusion in threshold models.

Chapter 5

Hybrid Logic for Diffusion

5.1 Introduction

This chapter, based on work published in [63], designs a general logic for reasoning about complex diffusion processes within social networks.

To do so, we build on work presented earlier. Conceptually, the idea of equipping agents with several properties taking different values, originates from the work from [62] presented in the previous chapter, generalizing the "one-property" opinion dynamics of [120]. However, our discussion takes a significant turn here: from the case-studies of specific social phenomena presented earlier, we now jump into a more general and more technical discussion. Indeed, while Chapter 4 restricted the modeling of (opinion) diffusion to *two* properties of agents, the framework to be introduced here allows us to reason about a wide class of complex phenomena involving an *unbounded* number of interacting properties of agents.

Formally, our framework is a dynamic extension of standard hybrid logic [10] with a special kind of atomic propositions to describe the agents' properties. The axiomatization of the static fragment of our logic is very similar to the axiomatizations of [54, 10] and the one of our full dynamic logic borrows the reduction technique from Dynamic Epistemic Logic [24, 70].

5.1.1 Outline

This chapter is structured as follows. Section 5.1.2 recalls the type of diffusion phenomena that our framework is designed to capture: local "contagion" processes, as they are typically considered within social network analysis. Section 5.2 defines our logic for diffusion in social networks. The static part of the logic is introduced first (5.2.1) and *dynamic transformations* are then added to it (5.2.2) to obtain the full dynamic setting. Section 5.3 provides a complete axiomatization of the logic. Section 5.4 shows how the logic applies to documented examples of network behaviors: we first consider an example of diffusion of micro-finance loans in villages (5.4.2), and we then go back to the phenomenon of pluralistic ignorance (5.4.1) discussed in the previous chapter. Finally, section 5.5 discusses possible extensions of the framework and summarizes the content of the chapter.

5.1.2 Complex Diffusion Phenomena

We are interested in a specific type of processes: the threshold-limited diffusion phenomena typically considered in network analysis, as introduced in Section 2.1.2.

Let us briefly recall here the most relevant features of the simple SI (susceptible/infected) epidemics example introduced in 2.1.2. The way an infection spreads within a community depends on how contagious the disease is, on how many agents are currently infected, on their positions in the network, and on the size and structure of the network. For instance, in the simplest version of such an SI example, the diffusion dynamics is driven by the following contagion rule: If any of your neighbors is infected, become infected yourself at the next moment, and stay infected forever. Given such an assumption about contagion, and a finite connected population, it is easy to see that all agents will end up being infected forever after a finite amount of time steps, which is bounded by the diameter of the graph.

Different contagion dynamics can lead to to differences in the long term dynamics: As mentioned in Section 2.1.2, we can also assume that after being infected, an agent could immediately recover and become susceptible again. According to this contagion rule, agents might find themselves alternating forever between being infected and being susceptible and the community will never reach a stable state with respect to the epidemics.

In this simple SI example, what determines the future health status of agents is their current health status and the one of their neighbors. That is, only one property (the health status) of the agents is relevant, and this property can take only two values (SI). But things are often much more complex. First, the health status could count more possible values (think of the three values SIR version, where agents recover and stop being contagious after being infected). Diffusion phenomena typically involve several properties of agents. Some other features of agents which might interfere with the diffusion of the infection. For instance, imagine a genetic type such that agents of this type are immune to the disease or stay infected for longer. In this case, the epidemic behavior reveals more complexity, and the diffusion rule needs to combine several properties of agents to take into account all factors. Some of these properties might be spreading within the network, under various contagion rules, and some might not be spreading at all.

Two aspects of the above examples will be particularly relevant to this chapter, for which we fix some vocabulary to prevent any confusion. First, agents have certain *properties* such as health status, genetic type, age, gender, hair color, etc, some of which might be spreading within the network, under various rules. For each agent all these properties are instantiated by particular *features* (or *values*), such as infected, is of the immune genetic type, 34 years old, female, redhead, etc. The features of some of these properties are spreading within the network (in our simple examples, only the health status features are). For each property, the associated possible features will come from some fixed set of values, such as: the three possible health states, numbers 1 to, say, 130, male or female, a set of possible hair colors, etc. We will represent this assignment of one value to each property of agents by a particular kind of atomic propositions and a constraint in the valuation in the framework developed in the next section.

The second thing to remark is that the dynamics are defined in a purely *local* way. In the above, an agent changes her health status from being susceptible to being infected if at least one of her neighbors is infected. Other kinds of local dynamics could be considered: for instance, dying your hair red if all of your friends have red hair or if at least one of your friends has a friend who has red hair. This type of local conditions is ideally described by formulas of a modal language. Thus, using an extension of basic modal logic will provide a natural way of defining a large variety of dynamic processes on social networks.

We want a framework to capture the laws of complex dynamic phenomena. More concretely, we design a general logic to model the change of repartition of properties within social networks, which can accommodate: 1) as many properties as wanted 2) as many values as wanted for each of those properties 3) any transformation rule which can be defined in the logic in terms of those properties, while keeping the network structure (and the agents' names) the same. This is what we are going to propose in this chapter.

5.2 Logic for Diffusion in Social Networks

In this section, we introduce a hybrid logical framework to reason about the change of distribution of features among agents within social networks. We start with the static part of the logic, which we call *logic for social networks*, to model the situation of agents in the network at a given moment. We then move on to the full dynamic *logic for diffusion in social networks* to represent the evolution of such situations.

5.2.1 Static Logic for Social Networks (N)

As in [145], our framework includes the standard tools from hybrid logic [10] to be able to talk about the network structure. Hence, following [145], our formulas will have an indexical reading.

The main novelty in our static logic is the format of the atomic propositions used to talk about the properties of the agents. Recall that we view agents as having certain features that are instantiations of some fixed properties under consideration. Instead of a set of standard propositional variables, we use equational statements to talk about features of agents. We assume that each agent has n different relevant properties, to each of which is assigned one value from a finite set. To avoid any later confusion, let us first remind our reader of the vocabulary we will be using: we use the term "property" to refer to for instance age, gender, health status, etc., and the term "feature" to refer to the value assigned to such a property, for instance 34 years old, infected, redhead, etc. In this sense, a property is a "feature variable" and a feature is a value taken by this variable.

More formally, we fix a finite set of *feature variables* $\{V_1, V_2, ..., V_n\}$ representing n different properties of agents, where each variable V_l is associated with a given finite *value set* R_l. The atomic propositions (or feature propositions) of our language will then be defined in the following way:

5.2.1. Definition. [Feature propositions] A *feature proposition* is of the form

$$V_l = r,$$

for some $l \in \{1, \dots, n\}$ and some $r \in R_l$. The set of all feature propositions (for fixed sets of variables and values) will be denoted FP.

The intuition is that the proposition $V_l = r$ is true of an agent if and only if the agent possesses feature r of property V_l. For instance, assuming that we have two properties V_g for gender and V_h for health status, we could write $V_g = f$ to express that an agent is female and $V_h = i$ to express that that an agent is infected.[1]

In addition to the finite set of feature propositions (FP), we will assume a countable infinite set of nominals (NOM) used as names for agents in networks, just as nominals are used to refer to possible states in traditional hybrid logic [10]. We can now give the syntax for the static social networks language:

5.2.2. Definition. [Syntax for social networks language $\mathcal{L}_{\mathcal{N}}$] The syntax of the social networks language, denoted $\mathcal{L}_{\mathcal{N}}$, is given by:

$$\varphi \ ::= \ V_l = r \ \mid \ i \ \mid \ \neg\varphi \ \mid \ \varphi \wedge \varphi \ \mid \ F\varphi \ \mid \ U\varphi \ \mid \ @_i\varphi \ ,$$

where $V_l = r \in \mathsf{FP}$ and $i \in \mathsf{NOM}$.[2]

[1] Note that feature propositions can be viewed as a generalization of classical propositional variables. Given a classical propositional variable P one can add a variable V_P and let $R_P = \{1, 0\}$. Then $V_P = 1$ will represent that P is true and $V_P = 0$ will represent that P is false (i.e. $\neg P$).

[2] We will use the abbreviations $\vee$, $\rightarrow$, and $\leftrightarrow$ in the standard way and denote the dual operator of F by $\langle F\rangle$ and the dual of U by $\langle U\rangle$. Moreover, we define: $\bigwedge_{i=1}^{n} \varphi_i := \bigwedge_{\varphi\in\{\varphi_1,\dots,\varphi_n\}} \varphi := (...((\varphi_1 \wedge \varphi_2) \wedge \varphi_3) \wedge ...) \wedge \varphi_n)$, and $\bigvee_{i=1}^{n} \varphi_i := \bigvee_{\varphi\in\{\varphi_1,\dots,\varphi_n\}} := (...((\varphi_1 \vee \varphi_2) \vee \varphi_3) \vee ...) \vee \varphi_n)$.

The intuitive meaning of a formula $F\varphi$ is that "φ is true of all my network-neighbors" and the intuitive meaning of a formula $@_i\varphi$ is that "φ is true of the agent named i" – note the indexical reading of formulas here! The U-operator is the global modality quantifying over all agents in the network and $U\varphi$ is read as "φ is true of all agents in the network".

As previously mentioned, we have assumed that a fixed set of feature propositions FP is given. This assumption will be made throughout the rest of the chapter unless otherwise specified. Whenever we need to be explicit about the set of feature propositions to which our language is relative, we will use the notation $\mathcal{L}_{\mathcal{N}}(\mathsf{FP})$.

Before defining the semantics, let us first introduce the notion of assignment and define our models. Intuitively, an assignment will assign specific values to the set of variables, hence determining the features of a given agent:

5.2.3. Definition. [Assignment/full assignment] An *assignment* (or *partial assignment*) is a partial function s from $\{1, ..., n\}$ to $\bigcup_{l=1}^{n} R_l$ such that $s(l) \in R_l$ for all $1 \leq l \leq n$ where s is defined. The set of all assignments is denoted by $\mathcal{V}$. For a given assignment s, the *domain* of s is denoted by $dom(s)$. A *full assignment* is an assignment s such that $dom(s) = \{1, ..., n\}$. The set of all full assignments is denoted $\mathcal{V}^{full}$.

The idea is that an assignment s assigns a feature $s(l) \in R_l$ to the feature variable V_l, for each $l \in dom(s)$. Thus, a *full* assignment s assigns a feature $s(l)$ to *every* feature variable V_l ($l \in \{1, ..., n\}$).[3]

5.2.4. Definition. [Network model] A *network model* is a tuple $\mathcal{M} = (A, \asymp, g, \nu)$, where: A is a non-empty set of agents, $\asymp$ is a binary relation on A representing the network structure, $g : \mathsf{NOM} \to A$ is a function assigning an agent to each nominal, and $\nu : A \to \mathcal{V}^{full}$ is a valuation assigning a full assignment $\nu(a)$ to each agent $a \in A$, i.e. a complete specification of the features of each agent in the network. The pair $(A, \asymp)$ will be referred to as a *frame* and a model built on a frame $(A, \asymp)$ is simply a model obtained by adding a g and a ν to the frame.

For instance, if we have two properties or "feature variables", health status and gender, a full assignment assigns one value for each variable to each agent in the network. In other words, no property of any agent is left undefined in a model. We can move on to the semantics for the language $\mathcal{L}_{\mathcal{N}}$:

[3]The reader might notice that we only need *full* assignments to define models. However, partial assignments will simplify things when we will define our dynamic logic for diffusion in social networks in Section 5.2.2.

5.2.5. Definition. [Semantics of $\mathcal{L}_\mathcal{N}$] Given a $\mathcal{M} = (A, \asymp, g, \nu)$, an $a \in A$ and a formula $\varphi \in \mathcal{L}_\mathcal{N}$, we define the truth of φ at a in $\mathcal{M}$ inductively by:

$$
\begin{array}{lll}
\mathcal{M}, a \models V_l = r & \text{iff} & \nu(a)(l) = r \\
\mathcal{M}, a \models i & \text{iff} & g(i) = a \\
\mathcal{M}, a \models \neg\varphi & \text{iff} & \text{it is not the case that } \mathcal{M}, a \models \varphi \\
\mathcal{M}, a \models \varphi \wedge \psi & \text{iff} & \mathcal{M}, a \models \varphi \text{ and } \mathcal{M}, a \models \psi \\
\mathcal{M}, a \models U\varphi & \text{iff} & \text{for all } b \in A; \mathcal{M}, b \models \varphi \\
\mathcal{M}, a \models F\varphi & \text{iff} & \text{for all } b \in A; a \asymp b \text{ implies } \mathcal{M}, b \models \varphi \\
\mathcal{M}, a \models @_i\varphi & \text{iff} & \mathcal{M}, g(i) \models \varphi
\end{array}
$$

We say that a formula φ is *satisfiable* if there is a model $\mathcal{M} = (A, \asymp, g, \nu)$ and an agent $a \in A$ such that $\mathcal{M}, a \models \varphi$ (and unsatisfiable otherwise). If this is the case, we will also simply say that a satisfies φ (taking $\mathcal{M}$ to be given). Two formulas φ and ψ are said to be *pairwise unsatisfiable* if $\varphi \wedge \psi$ is unsatisfiable. Given a model $\mathcal{M} = (A, \asymp, g, \nu)$ and a formula φ we write $\mathcal{M} \models \varphi$ if $\mathcal{M}, a \models \varphi$ for all $a \in A$. A formula φ is said to be *valid with respect to a class of frames* if $\mathcal{M} \models \varphi$ for all models $\mathcal{M}$ built on some frame from the class. A formula is said to be just *valid* if it is valid with respect to the class of all frames. The logic consisting of the set of all valid formulas will be denoted N and referred to as the logic for social networks.

While choosing a *modal* language seems natural to describe network structures, the reader might wonder why we choose a *hybrid* one. [38] has shown that some global properties of graphs standardly discussed in graph theory are neither definable in basic modal language (even if one adds a transitive closure modal operator to the language) nor in any bisimulation invariant extension of it, such as modal μ-calculus: connectivity, acyclicity, and Hamiltonian property (i.e., whether there is a cycle passing through each vertex of a graph exactly once), for instance. Add nominals and $@_i$ normal modal operators and all those properties become definable, as [38] shows. While our language does not include the transitive closure operator used in [38] and therefore cannot express connectivity and acyclicity with the same succinctness[4], it can express the Hamiltonian property in the exact same way as introduced in [38].

We leave the full expressivity comparison between the present framework and others for future research, but we give four examples of global properties of networks expressible in our language: irreflexivity (no agent is linked to itself), symmetry (if a first agent is linked to a second one, then the second one is also linked to the first one), full connectivity (all different agents are linked to each other), n-connectivity (there is a path of length at most n between any two pair of different agents):

[4]Section 5.5.2 discusses extensions of our framework with transitive closure operators (of both the static and dynamic modalities).

A frame $(A, \asymp)$ is

irreflexive iff $(A, \asymp) \models @_i \neg \langle F \rangle i$,

symmetric iff $(A, \asymp) \models @_i \langle F \rangle j \to @_j \langle F \rangle i$,

fully connected iff $(A, \asymp) \models @_i (\neg j \to \langle F \rangle j)$,

n-connected $(n \in \mathbb{N} \geq 1)$ iff $(A, \asymp) \models @_i (\neg j \to (\langle F \rangle j \vee \langle F \rangle^2 j \vee .... \langle F \rangle^n j)$, where $\langle F \rangle^k$ is given by:

$$\langle F \rangle^1 := \langle F \rangle$$
$$\langle F \rangle^{k+1} := \langle F \rangle \langle F \rangle^k.$$

All such global connectivity properties will affect the diffusion phenomena we are interested in. For instance, recall the initial epidemic "SI model" example from Section 5.1: if a network is n-connected, all agents will become infected after a period of length at most n. Hence, the hybrid static framework introduced so far is rather promising: it is expressive enough to describe the distribution of features among agents in the network, as well as some network properties which are particularly relevant to how those features will be redistributed in the future. In the next subsection, we introduce the tools to talk precisely about *changes* or *transformations* of such network situations.

5.2.2 Dynamic Logic for Diffusion in Networks (ND)

We now extend our logic to deal with the dynamics of networks. We make two important design choices. First, we will be concerned exclusively with one particular type of change: the change of distribution of features of agents within a social network structure. This means that we assume that agents do not change names and that the network structure is fixed. Second, we take a very general point of view: our agents are essentially just bundles of features with stable names, and the question of *how* such features should change is so open-ended that we consider that the safest option is to offer a framework which is general enough to allow for any such type of change, as long as it is locally definable in terms of our language. Consequently, our setting can be refined in many ways to accommodate different types of applications and represent their corresponding dynamics. Our framework allows to "plug-in": 1) how many properties of agents are relevant, 2) how many values each of these properties can take and 3) according to which rules such static models should be updated, i.e, how those features should be redistributed on the network. In other words, we are abstracting as much as possible from particular diffusion examples given by the networks analysis literature by building a framework which can deal with most of them.

What we want is a way to obtain a new model from a given model through some transformation. In this respect, our dynamic modalities will be comparable to the modalities for event models of Dynamic Epistemic Logic [24, 70]. However, instead of event models, we will talk about *dynamic transformations*[5].

5.2.6. Definition. [Dynamic transformations] A *dynamic transformation* is a pair $\mathcal{D} = (\Phi, \mathsf{post})$ consisting of a non-empty finite set Φ of pairwise unsatisfiable formulas (from the language $\mathcal{L}_{\mathcal{ND}}$ to be defined in Definition 5.2.7)[6] and a post-condition function $\mathsf{post} : \Phi \to \mathcal{V}$. The set Φ will be referred to as "preconditions", and given a precondition $\varphi \in \Phi$, we will call the assignment $\mathsf{post}(\varphi) \in \mathcal{V}$ the "post-condition" of φ.

Note that the post-conditions are partial assignments and not full assignments. The intuition behind this definition is that if an agent satisfies some formula $\varphi \in \Phi$ (in which case, φ is necessarily unique), then after the dynamic transformation $\mathcal{D}$, a changes her features as specified by $\mathsf{post}(\varphi)$. As $\mathsf{post}(\varphi)$ is a partial assignment a does not change all her features, only the ones in $dom(\mathsf{post}(\varphi))$.

On the syntactic level, we add formulas of the form $[\mathcal{D}]\varphi$ for a given dynamic transformation $\mathcal{D}$. In the following we consider a fixed set of dynamic transformations to be given and denote it by DT. Here is the syntax of our full dynamic language:

5.2.7. Definition. [Syntax of language for diffusion in social networks $\mathcal{L}_{\mathcal{ND}}$] The syntax of the language for diffusion in social networks, denoted $\mathcal{L}_{\mathcal{ND}}$, is given by:

$$\varphi \;::=\; V_l = r \mid i \mid \neg\varphi \mid \varphi \wedge \varphi \mid F\varphi \mid U\varphi \mid @_i\varphi \mid [\mathcal{D}]\varphi \,,$$

where $V_l = r \in \mathsf{FP}$, $i \in \mathsf{NOM}$, and $\mathcal{D} \in \mathsf{DT}$.

As for the static language $\mathcal{L}_{\mathcal{N}}$, whenever needed, to make explicit which set of dynamic transformations DT (and feature propositions FP) a language is built upon, we use the notation $\mathcal{L}_{\mathcal{ND}}(\mathsf{DT})$ ($\mathcal{L}_{\mathcal{ND}}(\mathsf{FP}, \mathsf{DT})$).

The satisfaction of formulas involving dynamic modalities relies on transforming the model at hand. This is captured by the following definition:

[5] The dynamic transformations in this chapter are defined by making use of both pre- and postconditions. This is in line with the work done in Dynamic Epistemic Logic with "fact change" and epistemic planning, see e.g. [55]. An alternative way to encode atomic fact change doesn't rely on postconditions but uses a flip-operator which changes the valuation of an atomic sentence to its opposite value. The latter approach is developed in [17].

[6] We have to be a little careful here. To avoid circular definitions we cannot allow the dynamic transformation $\mathcal{D} = (\Phi, \mathsf{post})$ to have precondition formulas in Φ involving $\mathcal{D}$ itself. Nevertheless, we can allow formulas of $\mathcal{L}_{\mathcal{ND}}$ in Φ constructed on an "earlier stage" in a simultaneous inductive definition of dynamic transformations and the language $\mathcal{L}_{\mathcal{ND}}$. In other words, one should view Definition 5.2.6 and Definition 5.2.7 as one simultaneous recursive definition. The issue is similar to the issue of defining the full language of Dynamic Epistemic Logic [70, Ch. 6].

5.2.8. Definition. [Transformation updates] Given a model $\mathcal{M} = (A, \asymp, g, \nu)$ and a dynamic transformation $\mathcal{D} = (\Phi, \mathsf{post})$, the *updated* model under the transformation $\mathcal{D}$ is $\mathcal{M}^{\mathcal{D}} = (A, \asymp, g, \nu')$, where ν' is defined by:

$$\nu'(a)(l) \quad = \quad \begin{cases} \mathsf{post}(\varphi)(l) & \text{if there is a } \varphi \in \Phi \text{ such that } \mathcal{M}, a \models \varphi \\ & \text{ and } l \in dom(\mathsf{post}(\varphi)) \\ \nu(a)(l) & \text{otherwise} \end{cases} \tag{5.1}$$

for all $a \in A$ and all $l \in \{1, ..., n\}$.

As previously mentioned, the intuition is that if an agent satisfies a $\varphi \in \Phi$ then, after the dynamic transformation $\mathcal{D}$, she changes her features as specified by $\mathsf{post}(\varphi)$. More formally, assume that an agent a satisfies φ and consider the variable V_l. If $l \notin dom(\mathsf{post}(\varphi))$, then $V_l = r$ will be true of a after $\mathcal{D}$ if, and only, if $V_l = r$ was true of a before $\mathcal{D}$. On the other hand, if $l \in dom(\mathsf{post}(\varphi))$, then $V_l = r$ will be true of a after $\mathcal{D}$ if, and only, if $\mathsf{post}(\varphi)(l) = r$. Note that the "otherwise" case in (5.1) takes care of two situations, namely the situation where there are no formulas in Φ true of the agent a, and the situation where there might be a formula $\varphi \in \Phi$ true of a, but the feature in question, l, is not in the domain of $\mathsf{post}(\varphi)$.

The semantics of the dynamic language can now be given:

5.2.9. Definition. [Semantics of $\mathcal{L}_{\mathcal{ND}}$] Given a $\mathcal{M} = (A, \asymp, g, \nu)$, an $a \in A$ and a formula $\varphi \in \mathcal{L}_{\mathcal{ND}}$, we define the truth of φ at a in $\mathcal{M}$ inductively as in Definition 5.2.5 with the additional clause:

$$\mathcal{M}, a \models [\mathcal{D}]\varphi \quad \text{iff} \quad \mathcal{M}^{\mathcal{D}}, a \models \varphi.$$

Satisfiability, validity, and pairwise unsatisfiability are generalized in the obvious way from Definition 5.2.5. The logic consisting of the set of all valid $\mathcal{L}_{\mathcal{ND}}$-formulas will be denoted ND and referred to as logic for diffusion in social networks.

Before moving on, let us consider the SIR (susceptible, infected, recovered) example introduced in Section 2.1.2 concerning diffusion of a disease again. Here we might have two variables V_{HS} and V_{GI} keeping track of the health status of the agents and whether they are genetically immune, i.e. R_{HS} = {susceptible, infected, recovered } and R_{GI} = {yes, no}. Thus, $V_{HS} = \mathit{susceptible} \wedge V_{GI} = \mathit{no}$ is true of an agent if she is susceptible to the disease and she is not genetically immune. We could then specify the following dynamic transformation $\mathcal{D} = (\Phi, \mathsf{post})$, for instance:

Φ :	post :
$V_{HS} = \mathit{susceptible} \wedge V_{GI} = \mathit{no} \wedge \langle F \rangle V_{HS} = \mathit{infected}$	$\mathsf{post}(HS) = \mathit{infected}$
$V_{GI} = \mathit{yes}$	$\mathsf{post}(HS) = \mathit{recovered}$
$V_{HS} = \mathit{infected}$	$\mathsf{post}(HS) = \mathit{recovered}$

This dynamic transformation represents the fact that a non genetically immune susceptible agent becomes infected if at least one of her neighbors is infected, while a genetically immune agent is immediately recovered and does not get infected ever. Moreover, after being infected an agent moves to being recovered. As there is no specification of how agents would move from being recovered to being susceptible (or infected), recovered agents become immune to the disease too. This example is minimal in the sense that it involves only two properties of agents and assumes that one property is spreading (health status) while the other is not (genetic immunity status). However, our framework can go way beyond this simple example as it allows for combining several properties, spreading or not. More complex examples of applications of ND can be found in Section 5.4.

It follows from Definition 5.2.8 that, for every network model $\mathcal{M}$ and every dynamic transformation $\mathcal{D}$, the updated network model $\mathcal{M}^{\mathcal{D}}$ always exists. Moreover, no agent from A is deleted when moving to the new model $\mathcal{M}^{\mathcal{D}}$. Thus for every pair $(\mathcal{M}, a)$ of a network model $\mathcal{M}$ and an agent a of $\mathcal{M}$, and for every dynamic transformation $\mathcal{D}$, the pair $(\mathcal{M}^{\mathcal{D}}, a)$ exists. Moreover, dynamic transformations are "functional", in the sense that each dynamic transformation $\mathcal{D}$ behaves as a function on the class of pointed network models $(\mathcal{M}, a)$, in accordance with most of the examples in the network analysis literature. This is reflected in the logic by the fact that all dynamic transformations are their own duals, i.e.

$$[\mathcal{D}]\varphi \leftrightarrow \neg[\mathcal{D}]\neg\varphi,$$

is a validity for all dynamic transformations $\mathcal{D}$ and all formulas φ. For these reasons, we can always define the sequential application of the same dynamic transformation $\mathcal{D}$ in a straightforward way as follows:

5.2.10. Definition. [$\mathcal{M}^{k\mathcal{D}}$] Given a network model $\mathcal{M}$ and a dynamic transformation $\mathcal{D}$, let $\mathcal{M}^{k\mathcal{D}}$ be defined recursively for all $k \in \mathbb{N}_0$ by:

$$\begin{array}{rcl} \mathcal{M}^{0\mathcal{D}} & := & \mathcal{M} \\ \mathcal{M}^{(k+1)\mathcal{D}} & := & (\mathcal{M}^{k\mathcal{D}})^{\mathcal{D}}. \end{array}$$

Some long-term behaviors of networks can be observed. Given a network model $\mathcal{M}$ and a dynamic transformation $\mathcal{D}$, an interesting question is whether the network stabilizes, that is, whether successive updates by $\mathcal{D}$ will result in a network model which does not change under further update by $\mathcal{D}$, i.e. a fixed-point of the model transformation $\mathcal{D}$.

5.2.11. Definition. [Stability of a model] A network model $\mathcal{M}$ is said to be *stable* under a dynamic transformation $\mathcal{D}$, if $\mathcal{M} = \mathcal{M}^{\mathcal{D}}$. $\mathcal{M}$ is said to *stabilize* under the dynamic transformation $\mathcal{D}$, if there is a $k \in \mathbb{N}_0$ such that $\mathcal{M}^{k\mathcal{D}}$ is stable.

Can our logic say something about such limit behaviors of networks? Yes, it can: it can capture the notion of stability. Let us explain how to express in our language that a network is stable.[7] Given a model $\mathcal{M} = (A, \asymp, g, \nu)$, the full assignment $\nu(a)$ completely describes the features of a, thus the complete features of a is expressed by:

$$\varphi_{\nu(a)} := \bigwedge_{l=1}^{n} V_l = \nu(a)(l).$$

Moreover, note that the set of all possible full assignments $\mathcal{V}$ is finite. Thus, we can "quantify" over it in our language and express that a network model is stable under $\mathcal{D}$ by[8]:

$$\varphi_{stable(\mathcal{D})} := \bigwedge_{s \in \mathcal{V}^{full}} \big(\varphi_s \to [\mathcal{D}]\varphi_s\big). \tag{5.2}$$

That this is in fact so follows from the following lemma:

5.2.12. Lemma. *A network model $\mathcal{M}$ is stable under $\mathcal{D}$ if, and only if,*

$$\mathcal{M} \models \varphi_{stable(\mathcal{D})}.$$

Let us summarize what we have done so far. First, we have defined a static logic to talk about features of agents in a social network. Then, we have defined the set of transformations of the distribution of those features which are locally definable in terms of preconditions and postconditions within our (restricted) language. Moreover, we have shown that our language can capture some static properties of networks, such as n-connectedness, and some dynamic properties of network models such as stability. In a nutshell, we have presented a logic able to describe the type of states of social networks and the type of changes which we wanted to capture. In the next section, we will consider what kind of *reasoning* about those networks is supported by our logic, by giving a complete proof-system for it.

[7]While our language can express stability, it cannot express stabilization. The straight forward way would be to add the Propositional Dynamic Logic (PDL) transitive closure construct $*$ to our modality $[\mathcal{D}]$ [95]. However, it would be interesting to find a formula without the $\langle \mathcal{D}^* \rangle$ operator that defines stabilizing networks, just as done in [166] for the case of a logic of preference change. We leave this for future research. Stabilization and the transitive closure operator $*$ are discussed in more details in the concluding Section 5.5.

[8]Another way of expressing that a network model is stable would be to follow the line of [120]. If $V_L = r$ is true of some agent and the network is stable, this means that none of the preconditions $\varphi \in \Phi$ of $\mathcal{D}$ for which $\mathsf{post}(\varphi)$ would change the value of V_l can be satisfied at the agent. Then, for every feature we can write the conjunction of the negation of all preconditions that would change this feature. Finally, we can take the disjunction over all possible features and thereby obtain a formula for a network being stable. This, of course, would result in a much more complex formula, however, it would avoid the explicit use of the $[\mathcal{D}]$ modality.

5.3 Axiomatization

In this section, we will provide sound and complete Hilbert-style proof systems for the logics of Section 5.2. The axiomatization of the static logic N follows that of [54, 10] with a few modifications, while the axiomatization of the dynamic logic ND expands that of the static logic with "reduction axioms" – a standard technique of Dynamic Epistemic Logic [24, 70].

Before giving the Hilbert-style axiomatization of N, we recall some standard terminology for Hilbert-style proof systems: A proof of φ is a finite sequence of formulas ending with φ such that every formula in the sequence is either an axiom or follows from previous formulas in the sequence using one of the proof rules. We denote this by $\vdash \varphi$. We use $\vdash_{\mathsf{S}}$ for provable in the proof-system for N and $\vdash_{\mathsf{D}}$ for provable in the proof-system for ND. In the following, X will thus stand for either S or D. For a set of formulas Γ, $\Gamma \vdash_{\mathsf{X}} \varphi$ holds if there are $\psi_1, ..., \psi_n \in \Gamma$ such that $\vdash_{\mathsf{X}} \psi_1 \wedge ... \wedge \psi_n \to \varphi$. Given a set of formulas Σ, let $\mathsf{X} + \Sigma$ denote the logic obtained by adding all the formulas in Σ as axioms. That φ is provable in the logic $\mathsf{X} + \Sigma$ will then be denoted by $\vdash_{\mathsf{X}+\Sigma} \varphi$. A set of formulas Γ is said to be $\mathsf{X} + \Sigma$-inconsistent if $\Gamma \vdash_{\mathsf{X}+\Sigma} \bot$, and $\mathsf{X} + \Sigma$-consistent otherwise. A formula φ is pure if it does not contain any feature propositions. A set of formulas Σ is called substitution-closed if it is closed under uniform substitution of nominals by nominals.

5.3.1 Complete Axiomatization of N

The Hilbert-style axiomatization of N is shown in Figure 5.1. As previously mentioned, the axiomatization is fairly standard in the hybrid logic literature except for the axioms Char.Prop.1 and Char.Prop.2. While Char.Prop.1 ensures every variable V_l is assigned at least one value, Char.Prop.2 ensures that no variable V_l is assigned more than one value.

Soundness and completeness of this type of axiomatization are also standard results in the hybrid logic literature (see [54, 10]). Thus, we do not include the proofs of these properties and only state the completeness theorem here:

5.3.1. Theorem (Completeness of N). *Let Σ be a substitution-closed set of pure formulas. Every set of formulas that is $\mathsf{N} + \Sigma$-consistent is satisfiable in a model whose underlying frame validates all the formulas in Σ.*

This form of completeness theorem is typical for hybrid logic and highlights one of its advantages. As traditional with completeness proofs, when assuming that a formula φ is valid and that it is not provable, the set $\{\neg\varphi\}$ becomes consistent. Then, by the above theorem, we can derive a counter-model to φ, which yields a contradiction to the assumption that φ was not valid – thus φ must be provable. The benefit of the hybrid logic is that we can make sure

Axioms:	
All substitution instances of propositional tautologies	
$\bigwedge_{l=1}^{n}\big(\bigvee_{r\in R_l} V_l = r\big)$	Char.Prop.1
$\bigwedge_{l=1}^{n}\bigwedge_{r\in R_l}\big(V_l = r \to \bigwedge_{s\in R_l\setminus\{r\}}\neg V_l = s\big)$	Char.Prop.2
$X(\varphi\to\psi)\to(X\varphi\to X\psi)$[1]	K_X
$@_i(\varphi\to\psi)\to(@_i\varphi\to @_i\psi)$	$\mathrm{K}_@$
$@_i\varphi\leftrightarrow\neg @_i\neg\varphi$	$\mathrm{Selfdual}_@$
$@_i i$	$\mathrm{Ref}_@$
$@_i @_j\varphi\leftrightarrow @_j\varphi$	Agree
$i\to(\varphi\leftrightarrow @_i\varphi)$	Introduction
$\langle X\rangle @_i\varphi\to @_i\varphi$[1]	Back
$(@_i\langle X\rangle j\wedge @_j\varphi)\to @_i\langle X\rangle\varphi$[1]	Bridge
$\langle U\rangle i$	GM
Rules:	
From φ and $\varphi\to\psi$, infer ψ	Modus ponens
From φ, infer $X\varphi$[1]	Necessitation of X
From φ, infer $@_i\varphi$	Necessitation of @
From $@_i\varphi$, where i does not occur in φ, infer φ	Name
From $(@_i\langle X\rangle j\wedge @_j\varphi)\to\psi$, where $i\neq j$ and j does not occur in φ or ψ, infer $@_i\langle X\rangle\varphi\to\psi$[1]	Paste

[1] Here X denotes either F or U.

Figure 5.1: The Hilbert-style proof system of N.

that this counter-model is of a typical kind, namely a model where each world is named by a nominal. This further implies that the underlying frame validates all formulas in Σ (as Σ is substitution closed). Thus, our counter-model is based on a frame from the class of frames defined by Σ and we thereby automatically achieve completeness with respect to the class of frames defined by Σ. (Going through an extension of the standard translation of modal logic into first-order logic, one can see that pure formulas will always define first-order properties of frames [53].)

The strength of this kind of automatic completeness is easily illustrated by considering a particular class of networks, namely networks where the relation $\asymp$ is irreflexive and symmetric – corresponding to the undirected networks very often studied in social networks analysis.[9] As mentioned in Section 5.2.1 irreflexivity and symmetry can be expressed (by pure formulas) in N. Thus, adding all substitution instances of these formulas allows us to derive a complete axiomatization of the logic of undirected networks.

[9] This is also the class of networks, representing “friendship”, which is considered by [145, 120].

Axioms:
All axioms for N of Figure 5.1

$[\mathcal{D}]V_l = r \leftrightarrow \big(\bigvee_{\varphi\in\Phi,\ \mathsf{post}(\varphi)(l)=r}\varphi\big) \vee \big(\neg(\bigvee_{\varphi\in\Phi,\ l\in dom(\mathsf{post}(\varphi))}\varphi) \wedge V_l = r\big)$	Red.Ax.Prop.
$[\mathcal{D}]i \leftrightarrow i$	Red.Ax.Nom.
$[\mathcal{D}](\varphi \wedge \psi) \leftrightarrow [\mathcal{D}]\varphi \wedge [\mathcal{D}]\psi$	Red.Ax.$\wedge$
$[\mathcal{D}]\neg\varphi \leftrightarrow \neg[\mathcal{D}]\varphi$	Red.Ax.$\neg$
$[\mathcal{D}]@_i\varphi \leftrightarrow @_i[\mathcal{D}]\varphi$	Red.Ax.@
$[\mathcal{D}]F\varphi \leftrightarrow F[\mathcal{D}]\varphi$	Red.Ax.F
$[\mathcal{D}]U\varphi \leftrightarrow U[\mathcal{D}]\varphi$	Red.Ax.U
$[\mathcal{D}][\mathcal{D}']\varphi \leftrightarrow [(\mathcal{D};\mathcal{D}')]\varphi$	Red.Ax.DD

Rules:
All the rules for N of Figure 5.1

For all dynamic transformations $\mathcal{D}, \mathcal{D}' \in \mathsf{DT}$.

Figure 5.2: The Hilbert-style proof system of ND.

5.3.2 Complete Axiomatization of ND

We now move on to give a complete axiomatization of the full dynamic logic ND. The axiomatization is shown in Figure 5.2. The new axioms, referred to as *reduction axioms*, allow us to reduce all talk about dynamic properties of the network models to talk about their static properties. Moreover, they give us a better understanding of the dynamic transformations. For instance, the intuition behind the first reduction axiom Red.Ax.Prop. is that if the variable V_l is assigned the value r after the dynamic transformation $\mathcal{D}$, then before the transformation either *i*) one of the post-conditions of $\mathcal{D}$ that specify a change resulting in $V_l = r$, is satisfied, or *ii*) no precondition of $\mathcal{D}$ that specify a change to the value taken by variable V_l is satisfied and $V_l = r$ is already true.

The intuition behind the axiom Red.Ax.Nom. is that dynamic transformations do not change the names of agents. The axiom Red.Ax.$\wedge$ says that dynamic transformations commute with conjunction, while the axiom Red.Ax.$\neg$ says that they also commute with negation. That negation commutes with a dynamic modality might seem a bit surprising to readers familiar with public announcement logic or traditional dynamic epistemic logic. However, as dynamic transformations can always be executed (as discussed after Definition 5.2.9), $\neg\varphi$ is true after the dynamic transformation $\mathcal{D}$ if, and only if, it is not the case that φ is true after a dynamic transformation $\mathcal{D}$. The axioms Red.Ax.@, Red.Ax.F, and Red.Ax.U further state that the modalities $@_i$, F, and U commute with dynamic transformations. The fact that dynamic transformation modalities commute with the other modalities highlights the fact that dynamic transformations of network models can be reduced to local changes at each agent in the network models. As such, the reduction axioms provide new insights about the behavior of dynamic transformations.

The way we will show completeness of this proof system is the usual way in Dynamic Epistemic Logic, namely by providing a truth-preserving translation from ND into N. Before this, however, we need to define the composition of two dynamic transformations[10] as used in the last reduction axiom of Figure 5.2.[11]

5.3.2. Definition. [Composition of dynamic transformations] Given two dynamic transformations $\mathcal{D} = (\Phi, \mathsf{post})$ and $\mathcal{D}' = (\Phi', \mathsf{post}')$, the composition $(\mathcal{D}; \mathcal{D}') = (\Phi'', \mathsf{post}'')$ is such that

$$\begin{aligned} \Phi'' \;=\; & \{\varphi \wedge [\mathcal{D}]\psi \mid \varphi \in \Phi, \psi \in \Phi'\} \cup \{\varphi \wedge [\mathcal{D}]\big(\bigwedge_{\psi \in \Phi'} \neg\psi\big) \mid \varphi \in \Phi\} \\ & \cup \{\big(\bigwedge_{\varphi \in \Phi} \neg\varphi\big) \wedge [\mathcal{D}]\psi \mid \psi \in \Phi'\}, \end{aligned}$$

and post'' is such that

$$\begin{aligned} \mathsf{post}''(\varphi \wedge [\mathcal{D}]\psi)(l) &= \mathsf{post}'(\psi)(l)\ , \quad \text{if } l \in dom(\mathsf{post}'(\psi)) \\ \mathsf{post}''(\varphi \wedge [\mathcal{D}]\psi)(l) &= \mathsf{post}(\varphi)(l)\ , \quad \text{if } l \in dom(\mathsf{post}(\varphi)) \setminus dom(\mathsf{post}'(\psi)) \\ \mathsf{post}''(\varphi \wedge [\mathcal{D}]\big(\bigwedge_{\psi \in \Phi'} \neg\psi\big))(l) &= \mathsf{post}(\varphi)(l)\ , \quad \text{if } l \in dom(\mathsf{post}(\varphi)) \\ \mathsf{post}''(\big(\bigwedge_{\varphi \in \Phi} \neg\varphi\big) \wedge [\mathcal{D}]\psi)(l) &= \mathsf{post}'(\psi)(l)\ , \quad \text{if } l \in dom(\mathsf{post}'(\psi)). \end{aligned}$$

Note that this definition is well-defined as Φ'' will consist of pairwise unsatisfiable formulas. Moreover, while this definition might seem a bit complicated, this is only due to the fact that we have to take into account the following three cases for a given agent:

(i) One of the formulas in Φ is satisfied at the agent and afterwards the agent satisfies one of the formulas in Φ'.

(ii) One of the formulas in Φ is satisfied at the agent, but after the dynamic transformation $\mathcal{D}$ the agent does not satisfy any formula in Φ'.

[10] Note that we treat ";" purely as a semantic operation on dynamic transformations. One could also have included ";" directly in the syntax in Definition 5.2.7. However, we have chosen not to do so, as we only use this composition of dynamic transformations for proving completeness.

[11] At first sight the last reduction axiom Red.Ax.DD might seem superfluous. However, as we define our translation from ND to N "outside-in" on formulas, we need this reduction axiom for composition. If one defines the translation "inside-out" on formulas one would instead need "replacement of equivalents". However, "replacement of equivalents" cannot be derived from the other axioms in standard axiomatizations of public announcement logic and we suspect it cannot be here either. For an excellent discussion of these subtle issues concerning axiomatizations of dynamic epistemic logics see [160].

(iii) None of the formulas in Φ is satisfied at an agent, but after the dynamic transformation $\mathcal{D}$ the agent does satisfy one of the formula in Φ'.

These three cases give rise to the three sets in the definition of Φ''. Moreover, these three cases give rise to different definitions of post''. In the case of (i), there are three additional sub-cases according to whether a) the partial assignment $\mathsf{post}'(\psi)$ specifies a change of a feature, or b) $\mathsf{post}'(\psi)$ does not specify a change of a feature, but $\mathsf{post}(\varphi)$ does, or c) neither a) nor b) is the case. The cases a) and b) are directly taken care of in the definition of post'', whereas c) is indirectly taken care of by the fact that post'' might be partial assignment.

The following lemma, about composition of dynamic transformations will be useful:

5.3.3. LEMMA. *For every network model $\mathcal{M}$ and any two dynamic transformations $\mathcal{D}$ and $\mathcal{D}'$ we have that:*

$$(\mathcal{M}^{\mathcal{D}})^{\mathcal{D}'} = \mathcal{M}^{(\mathcal{D};\mathcal{D}')} \tag{5.3}$$

Before proving completeness we need to check the soundness of the reduction axioms (which, of course, also gives us soundness of the proof system). This is ensured by the following lemma:

5.3.4. LEMMA. *For all models $\mathcal{M} = (A, \asymp, g, \nu)$ and all $a \in A$, the following hold:*

$$\mathcal{M}, a \models [\mathcal{D}]V_l{=}r \quad \textit{iff} \tag{5.4}$$

$$\mathcal{M}, a \models \Big(\bigvee_{\varphi\in\Phi,\ \mathsf{post}(\varphi)(l)=r} \varphi\Big) \vee \Big(\neg\big(\bigvee_{\varphi\in\Phi,\ l\in dom(\mathsf{post}(\varphi))} \varphi\big) \wedge V_l{=}r\Big)$$

$$\mathcal{M}, a \models [\mathcal{D}]i \quad \textit{iff} \quad \mathcal{M}, a \models i \tag{5.5}$$

$$\mathcal{M}, a \models [\mathcal{D}]\neg\varphi \quad \textit{iff} \quad \mathcal{M}, a \models \neg[\mathcal{D}]\varphi \tag{5.6}$$

$$\mathcal{M}, a \models [\mathcal{D}](\varphi \wedge \psi) \quad \textit{iff} \quad \mathcal{M}, a \models [\mathcal{D}]\varphi \wedge [\mathcal{D}]\psi \tag{5.7}$$

$$\mathcal{M}, a \models [\mathcal{D}]@_i\varphi \quad \textit{iff} \quad \mathcal{M}, a \models @_i[\mathcal{D}]\varphi \tag{5.8}$$

$$\mathcal{M}, a \models [\mathcal{D}]F\varphi \quad \textit{iff} \quad \mathcal{M}, a \models F[\mathcal{D}]\varphi \tag{5.9}$$

$$\mathcal{M}, a \models [\mathcal{D}]U\varphi \quad \textit{iff} \quad \mathcal{M}, a \models U[\mathcal{D}]\varphi \tag{5.10}$$

$$\mathcal{M}, a \models [\mathcal{D}][\mathcal{D}']\varphi \quad \textit{iff} \quad \mathcal{M}, a \models [(\mathcal{D};\mathcal{D}')]\varphi \tag{5.11}$$

Proof: We only provide the proof of (5.4) and leave out the other cases. Let $\mathcal{M}^{\mathcal{D}}$ be $(A, \asymp, g, \nu')$, where ν' is defined as in (5.1). Then we have the following equivalences:

$$
\begin{array}{lcl}
t(V_l = r) & = & V_l = r \\
t([\mathcal{D}]V_l = r) & = & t\big(\big(\bigvee_{\varphi\in\Phi,\mathsf{post}(\varphi)(l)=r}\varphi\big) \vee \big(\neg(\bigvee_{\varphi\in\Phi,\ l\in dom(\mathsf{post}(\varphi))}\varphi) \wedge V_l{=}r\big)\big)
\end{array}
$$

$$
\begin{array}{lcllcl}
t(i) & = & i & t([\mathcal{D}]i) & = & t(i) \\
t(\neg\varphi) & = & \neg t(\varphi) & t([\mathcal{D}]\neg\varphi) & = & t(\neg[\mathcal{D}]\varphi) \\
t(\varphi\wedge\psi) & = & t(\varphi)\wedge t(\psi) & t([\mathcal{D}](\varphi\wedge\psi)) & = & t([\mathcal{D}]\varphi\wedge[\mathcal{D}]\psi) \\
t(\Box\varphi) & = & \Box t(\varphi)^1 & t([\mathcal{D}]\Box\varphi) & = & t(\Box[\mathcal{D}]\varphi)\ ^1 \\
 & & & t([\mathcal{D}][\mathcal{D}']\varphi) & = & t([(\mathcal{D};\mathcal{D}')]\varphi)
\end{array}
$$

[1] Here $\Box$ is either F, $@_i$, or U.

Figure 5.3: The translation $t : \mathcal{L}_{\mathcal{ND}} \to \mathcal{L}_{\mathcal{N}}$.

$$
\begin{array}{lll}
\mathcal{M}, a \models [\mathcal{D}]V_l = r & \text{iff} & \mathcal{M}^{\mathcal{D}}, a \models V_l = r \\
 & \text{iff} & \nu'(a)(l) = r.
\end{array}
$$

Note that, $\nu'(a)(l) = r$ is the case if, and only if, either there is a $\varphi \in \Phi$ such that $\mathcal{M}, a \models \varphi$ and $\mathsf{post}(\varphi)(l) = r$, or there is no such φ, but $\nu(a)(l) = r$. Now, the first disjunct of this disjunction is equivalent to $\mathcal{M}, a \models \big(\bigvee_{\varphi\in\Phi,\mathsf{post}(\varphi)(l)=r}\varphi\big)$ while the second is equivalent to $\mathcal{M}, a \models \big(\neg(\bigvee_{\varphi\in\Phi,\ l\in dom(\mathsf{post}(\varphi))}\varphi) \wedge V_l{=}r\big)$. Hence,

$$
\begin{array}{l}
\nu'(a)(l) = r \quad \text{iff} \\
\qquad \mathcal{M}, a \models \big(\bigvee_{\varphi\in\Phi,\mathsf{post}(\varphi)(l)=r}\varphi\big) \vee \big(\neg(\bigvee_{\varphi\in\Phi,\ l\in dom(\mathsf{post}(\varphi))}\varphi) \wedge V_l{=}r\big),
\end{array}
$$

and (5.4) has been proven. $\Box$

The soundness of the axiomatization of ND now follows from the soundness of the axiomatization of N together with Lemma 5.3.4. To show completeness we first define a translation t from $\mathcal{L}_{\mathcal{ND}}$ into $\mathcal{L}_{\mathcal{N}}$ as shown in Figure 5.3. Note that the translation t is not defined inductively on the usual notion of complexity of a formula. Therefore we cannot prove results regarding t by induction on this complexity. However, the complexity of the formula immediately succeeding a dynamic transformation decreases through the translation, and we can use this fact. A new complexity measure c, such that c decreases for every step of the translation, can be defined as follows:

5.3.5. DEFINITION. [New complexity measure c] Let the new complexity measure $c : \mathcal{L}_{\mathcal{ND}} \cup \mathsf{DT} \to \mathbb{N}$, be defined as follows:

$$
\begin{array}{lcl}
c(V_l = r) & = & 1 \\
c(i) & = & 1 \\
c(\neg\varphi) & = & 1 + c(\varphi) \\
c(\Box\varphi) & = & 1 + c(\varphi) \\
c(\varphi \wedge \psi) & = & 1 + max(c(\varphi), c(\psi)) \\
c([\mathcal{D}]\varphi) & = & (3 \cdot |\Phi| + 3 + c(\mathcal{D})) \cdot c(\varphi) \\
c(\mathcal{D}) & = & max\{c(\psi) \mid \psi \in \Phi\}
\end{array}
$$

where $\Box$ is "$@_i$", "F", or "U", and $\mathcal{D} = (\Phi, \mathsf{post})$.

We can show the following useful result: the translation of a dynamic formula can be reduced to the translation of a less complex formula:

5.3.6. LEMMA. *For all $i \in \mathsf{NOM}$, all $V_l = r \in \mathsf{FP}$, all $\varphi, \psi \in \mathcal{L}_{\mathcal{ND}}$, and all $\mathcal{D}, \mathcal{D}' \in \mathsf{DT}$ the following are true:*

1. $c([\mathcal{D}]i) > c(i)$
2. $c([\mathcal{D}]V_l = r) > c\big((\bigvee_{\varphi \in \Phi, \mathsf{post}(\varphi)(l) = r} \varphi) \vee (\neg(\bigvee_{\varphi \in \Phi,\ l \in dom(\mathsf{post}(\varphi))} \varphi) \wedge V_l = r)\big)$
3. $c([\mathcal{D}]\neg\varphi) > c(\neg[\mathcal{D}]\varphi)$
4. $c([\mathcal{D}]\Box\varphi) > c(\Box[\mathcal{D}]\varphi)$
5. $c([\mathcal{D}](\varphi \wedge \psi)) > c([\mathcal{D}]\varphi \wedge [\mathcal{D}]\psi)$
6. $c([\mathcal{D}][\mathcal{D}']\varphi) > c([\mathcal{D}; \mathcal{D}']\varphi)$

The proof of this lemma is quite cumbersome and involves tedious computation. Thus, for space reasons, we do not include it here. This lemma allows us to prove that every formula of the logic ND is provably equivalent to its translation in N:

5.3.7. LEMMA. *For all $\mathcal{L}_{\mathcal{ND}}$ formulas φ,*

$$\vdash_{\mathsf{D}} \varphi \leftrightarrow t(\varphi) \tag{5.12}$$

Proof: The proof goes by induction on the new c-complexity. For $c(\varphi) = 1$, φ is either of the form $V_l = r$ or of the form i. In both cases $\varphi = t(\varphi)$ and (5.12) is trivially satisfied. Now suppose that (5.12) holds for all φ with $c(\varphi) \leq n$. Then, we need to prove that (5.12) holds for all φ with $c(\varphi) = n+1$. Thus, assume that φ is a formula such that $c(\varphi) = n + 1$. We need to distinguish 4 cases: i) φ is of the form $\neg\psi$; ii) φ is of the form $\Box\psi$, with $\Box$ as in Def. 5.3.5; iii) φ is of the form $\psi_1 \wedge \psi_2$; and iv)φ is of the form $[\mathcal{D}]\psi$. We leave out the straightforward proofs of cases i),ii) and iii). To prove iv), we need to check the following sub-cases, corresponding to the 6 points of Lemma 5.3.6:

1. φ is of the form $[\mathcal{D}]i$. By Lemma 5.3.6.1 and induction hypothesis, $\vdash_{\mathsf{D}} i \leftrightarrow t(i)$. By Red.Ax.Nom and the fact that $t(i) = t([\mathcal{D}]i)$, $\vdash_{\mathsf{D}} [\mathcal{D}]i \leftrightarrow t([\mathcal{D}]i)$.

2. φ is of the form $[\mathcal{D}]V_l = r$. For readability, let us denote $(\bigvee_{\varphi\in\Phi,\mathsf{post}(\varphi)(l)=r} \varphi) \veebar (\neg(\bigvee_{\varphi\in\Phi,\ l\in dom(\mathsf{post}(\varphi))} \varphi) \wedge V_l = r)$ by χ. By Lemma 5.3.6.2 and induction hypothesis $\vdash_{\mathsf{D}} \chi \leftrightarrow t(\chi)$. By Red.Ax.prop and propositional logic, $\vdash_{\mathsf{D}} [\mathcal{D}]V_l = r \leftrightarrow t(\chi)$. Since $t([\mathcal{D}]V_l = r) = t(\chi)$, we conclude that $\vdash_{\mathsf{D}} [\mathcal{D}]V_l = r \leftrightarrow t([\mathcal{D}]V_l = r)$.

3. φ is of the form $[\mathcal{D}]\neg\psi$. By Lemma 5.3.6.3 and induction hypothesis $\vdash_{\mathsf{D}} \neg[\mathcal{D}]\psi \leftrightarrow t(\neg[\mathcal{D}]\psi)$. By Red.Ax$\neg$ and propositional logic, $\vdash_{\mathsf{D}} [\mathcal{D}]\neg\psi \leftrightarrow t(\neg[\mathcal{D}]\psi)$ and since $t([\mathcal{D}]\neg\psi) = t(\neg[\mathcal{D}]\psi)$, we can conclude that $\vdash_{\mathsf{D}} [\mathcal{D}]\neg\psi \leftrightarrow t([\mathcal{D}]\neg\psi)$.

4. φ is of the form $[\mathcal{D}]\Box\psi$. Similar to the case 3. just using Lemma 5.3.6.4 and the reduction axiom Red.Ax.$\Box$ instead.

5. φ is of the form $[\mathcal{D}](\psi_1 \wedge \psi_2)$. Similar to the case 3. just using Lemma 5.3.6.5 and and the reduction axiom Red.Ax.$\wedge$ instead.

6. φ is of the form $[\mathcal{D}][\mathcal{D}']\psi$. Similar to the case 3. just using Lemma 5.3.6.6 and and the reduction axiom Red.Ax.DD instead.

$\Box$

From lemma 5.3.7 and the soundness of the proof system, it follows directly that all formulas are also semantically equivalent to their translation:

5.3.8. LEMMA. *For all $\mathcal{L}_{\mathcal{ND}}$ formulas φ, all models $\mathcal{M} = (A, \asymp, g, \nu)$, and all $a \in A$,*

$$\mathcal{M}, a \models \varphi \iff \mathcal{M}, a \models t(\varphi)$$

Note that, translating pure formulas from $\mathcal{L}_{\mathcal{ND}}$ results in pure formulas in $\mathcal{L}_{\mathcal{N}}$. The general completeness result now follows:

5.3.9. THEOREM (COMPLETENESS FOR ND). *Let Σ be a substitution-closed set of pure $\mathcal{L}_{\mathcal{ND}}$-formulas. Every set of $\mathcal{L}_{\mathcal{ND}}$-formulas that is $\mathsf{D}+\Sigma$-consistent is satisfiable in a model whose underlying frame validates all the formulas in Σ.*

Proof: Assume that Γ is $\mathsf{D}+\Sigma$-consistent. For a set of $\mathcal{L}_{\mathcal{ND}}$-formulas X, let $t(X) := \{t(\varphi) \mid \varphi \in X\}$. Then $t(\Gamma)$ is $\mathsf{S}+t(\Sigma)$-consistent, for assume otherwise: Then there are $\varphi_1, ..., \varphi_n \in \Gamma$ such that $\vdash_{\mathsf{S}+t(\Sigma)} t(\varphi_1 \wedge ... \wedge \varphi_n) \to \bot$. But then also $\vdash_{\mathsf{D}+\Sigma} t(\varphi_1 \wedge ... \wedge \varphi_n) \to \bot$ (using lemma 5.3.7 on formulas in Σ) and by lemma 5.3.7, $\vdash_{\mathsf{D}+\Sigma} \varphi_1 \wedge ... \wedge \varphi_n \to \bot$, which is a contradiction to Γ being $\mathsf{D}+\Sigma$-consistent. Now by Theorem 5.3.1, $t(\Gamma)$ is satisfiable in a model $\mathcal{M}$ (which is also a model for $\mathcal{L}_{\mathcal{ND}}$), and by lemma 5.3.8 it follows that Γ is also satisfiable in $\mathcal{M}$.

Finally, for all pure formulas $\varphi \in \Sigma$, $t(\varphi)$ is a pure formula. Thus by Theorem 5.3.1 the underlying frame of $\mathcal{M}$ validates all of the formulas $t(\varphi) \in t(\Sigma)$. But by lemma 5.3.8 the underlying frame then also validates all $\varphi \in \Sigma$. $\square$

5.4 Applications

This section provides a few examples of the kind of modeling and reasoning about changes of distribution of features within social networks which ND allows for.

5.4.1 Pluralistic Ignorance

In Chapter 4, we have studied pluralistic ignorance from a dynamic perspective and we have discussed how the social network structure constrains the dynamics of its dissolution into an opposite situation where all agents express sincerely their private opinions. Our starting point was to note that such a phenomenon could not be modeled without distinguishing *two* properties of agents, their private belief state, which we call "inner belief" and their publicly observable behavior, which we call "expressed belief". As such, the phenomenon could not be captured by the "one property" (or "one-layer") framework for modeling belief change under conformity pressure offered by [120]. At the time, modeling cases of collective "failure" such as pluralistic ignorance was our first motivation for designing a framework allowing to model the change of *several* properties of agents, and hence for adopting the "multi-property" approach which we continued pursuing in this chapter. Our motivation here is much more general, since we now consider any set of features of agents changing under local influence. However, we briefly recall below how to model the case of pluralistic ignorance, as an example of application of our general framework to a well-known dynamic social phenomenon.

Let two variables V_I and V_E correspond to the properties of "inner belief" (private mental state) and "expressed belief" (observable behavior), respectively. Each variable takes values from the same set: $R_I = R_E = \{p, c, n\}$, where p represents belief in something (a *pro* opinion), c represents the belief in its negation (a *contra* opinion) and n represents the lack of belief or undecidedness (a neutral opinion).

Let us briefly recall the treatment of pluralistic ignorance proposed in Chapter 4. To model how a given opinion situation would evolve, we needed to assume some notion of social influence, that is, some dynamic transformation encoding how agents will change their belief states depending on the ones of their neighbors. One possibility, inspired by the (one-property) influence operator assumed in [120], was to consider that an agent is "brave enough" to express sincerely her actual private belief (i.e, $V_E = V_I$) at the next moment only when she has some "supporting" friend, i.e., some friends expressing what she privately believes or when she has no "conflicting" friends, i.e., no friend expressing a belief in the negation of what she privately believes.[12] Moreover, to reflect the intuition that influence affects, at least in good part, the observable side of agents, we considered that only their behavior (expressed opinion) was affected by social influence, not their private belief state. What was important for us was that their behavior depended on asymmetrical information: on the one hand, on what they themselves privately believe and, on the other hand, on what their neighbors publicly express.

We can now define formally the corresponding dynamic transformation as follows:[13]

$\mathcal{D}_E = (\Phi_E, \mathsf{post}_E)$:

$$\begin{aligned} \Phi_E = \quad & \{(V_I = p \wedge (\langle F \rangle V_E = p \vee [F]V_E = n)) \vee [F]V_E = p, \\ & (V_I = c \wedge (\langle F \rangle V_E = u \vee [F]V_E = n)) \vee [F]V_E = c, \\ & V_I = n \wedge \neg [F]V_E = p \wedge \neg [F]V_E = c\} \end{aligned}$$

$$\begin{aligned} \mathsf{post}_E\big((V_I = p \wedge (\langle F \rangle V_E = p \vee [F]V_E = n)) \vee [F]V_E = p\big)(V_E) = p \\ \mathsf{post}_E\big((V_I = c \wedge (\langle F \rangle V_E = c \vee [F]V_E = n)) \vee [F]V_E = c\big)(V_E) = c \\ \mathsf{post}_E\big(V_I = n \wedge \neg [F]V_E = p \wedge \neg [F]V_E = c\big)(V_E) = n \end{aligned}$$

Consider now a situation of pluralistic ignorance, in the sense that everybody privately believes something but expresses a belief in its negation, that is

[12] This correspondeds to agents of "type 1" from Chapter 4 as given by the Table 4.1, modulo the difference of notation.

[13] Note that this definition of the social influence operator is more succinct than listing all the possible preconditions and postconditions corresponding to all the cases listed in Table 4.1.

$\mathcal{M} \models U(V_I = p \wedge V_E = c)$. Now apply the transformation $\mathcal{D}_E$. Having another look at the preconditions set Φ_E above, note that none of them is satisfied at any agent. Therefore, none of the agents will change her behavior. As expected, a situation of pluralistic ignorance is stable under that dynamic transformation $\mathcal{D}_E$.

Now assume that the situation is slightly different: a unique agent, let it be named i, is expressing his private belief. This is what we have called a state of "unstable pluralistic ignorance": $\mathcal{M} \models @_i(V_I = p \wedge V_E = p) \wedge U\big(\neg i \rightarrow (V_I = p \wedge V_B = c)\big)$. It is easy to see that this situation is not stable under the transformation $\mathcal{D}_E$. For instance, considering the case of agent i itself: $\mathcal{M}, i \vDash [F]V_E = c$ and therefore $\mathcal{M}, i \vDash (V_I = c \wedge (\langle F\rangle V_E = c \vee [F]V_E = n)) \vee [F]V_E = c$. Since we know that $\mathsf{post}_E\big((V_I = c \wedge (\langle F\rangle V_E = c \vee [F]V_E = n)) \vee [F]V_E = c\big)(V_E) = c$, agent i will change his expressed belief state to a state in conflict with his private belief state, as a result of conformity pressure from all agents around him. But what about i's neighbors? Consider an arbitrary agent j such that $\mathcal{M}, i \vDash \langle F\rangle j$. Now $\mathcal{M}, j \vDash V_I = p \wedge \langle F\rangle V_E = p$ and therefore $\mathcal{M}, j \vDash (V_I = p \wedge (\langle F\rangle V_E = p \vee [F]V_E = n)) \vee [F]V_E = p$. Since we know that $\mathsf{post}_E\big((V_I = p \wedge (\langle F\rangle V_E = p \vee [F]V_E = n)) \vee [F]V_E = p\big)(V_E) = p$, agent j will now have an expressed belief state in agreement with his private state. And similarly for any neighbor of the initiator i. Hence, agent i and his neighbors have switched their expressed belief states after one application of the transformation. After one more step, i's friends' friends will express their actual inner state, and then i's friends' friends' friends, and so on. But then, by repeating the transformation n times, all agents at distance less or equal to n from i will have changed their state at least once.

This shows that the initial event of one agent being sincere will affect the entire population after some time. However, in the long run, nothing guarantees that process will stabilize. As we have shown in Proposition 4.3.2, stabilization depends on the network structure itself: if the network graph contains an odd cycle path, (that is, if the graph is not two-colorable), then a (connected, finite, symmetric and irreflexive) model in a state of unstable pluralistic ignorance will always stabilize and it will stabilize in a state where everybody expresses sincerely their private belief. This reflects properly the well-known fragility of pluralistic ignorance: one agent expressing her actual private belief state might influence everybody else into doing the same.[14]

5.4.2 Diffusion of Microfinance

The fact that social network structures affect the adoption of new technologies has been well-documented for some time already. A classical example is the

[14]For more details about the dissolution of pluralistic ignorance and a proof of the claim of stabilization, see Chapter 4. And for more details about stabilization of a sequence of models under iterated transformation rules, see Chapter 8.

diffusion of hybrid seed corn among Iowa farmers [140] (additionally, see the references in [35]). Still, the recent study [35] provides new insight about how social structures affect the spread of microfinance loans in small Indian villages. The authors of this study collected detailed data on various types of social ties and structures in 43 rural villages in Southern India before a microfinance institution entered the villages. Then, based on information from the microfinance institution, they compared the data on the social networks to the actual diffusion of microfinance loans in the villages.

It is argued in [35] that the diffusion of who is informed about the loaning possibilities is different from the diffusion of who chooses to participate in the microfinance loaning program. In the diffusion of microfinance, the most interesting parameter is who chooses to participate in the microfinance program. However, as shown by [35], this could not be estimated for individuals based on the participation of their neighbors in the social network. Moreover, the people who did not choose to participate in the microfinance program still passed on information about the program and thus, the diffusion of who was informed about the program did depend on whether an individual's neighbor was already informed (and chose to pass on the information). Hence, the two diffusion processes of information spreading and endorsement can come apart and as such the typical "SI Model" described in our introduction is not sufficient to represent such dynamics.

The spread of microfinance loans is a good example of why we might need *two* feature variables, one representing whether an individual is informed and one representing whether she has chosen to participate in the loaning program. The model presented in [35] is a probabilistic model and as such we cannot completely capture it in our framework. Nevertheless, we can describe some interesting variations. First, let us use two variables V_I and V_P, where V_I will keep track of who is informed about the microfinance program, and V_P will keep track of who has actually chosen to take part in the program. As value set we will assume that $R_I = R_P = \{y, n\}$ for "yes" and "no", with the obvious interpretation that an agent satisfies $V_I = y$ if she is informed about the program and that she satisfies $V_P = n$ if she is not participating in the program.

One could imagine that an agent becomes informed about the microfinance program as soon as one of her friends is either informed or has chosen to participate. However, [35] estimated that people participating in the program were much more likely to pass on information about it than non-participants. Still, the non-participants' passing on of the information could not be neglected either. Thus, an alternative principle could be that an agent becomes informed about the microfinance program if at least one of her friends is participating or all of her friends are already informed. This suggests the following dynamic transformation $\mathcal{D}_I = (\Phi_I, \mathsf{post}_I)$ of the diffusion of information about the program, where

$$\Phi_I = \{\langle F\rangle V_P = y \vee FV_I = y\}$$
$$\mathsf{post}_I\big(\langle F\rangle V_P = y \vee FV_I = y\big)(V_I) = y.$$

Concerning the diffusion of participation, [35] claims that, in their data at least, there is no endorsement effect and thus whether an agent chooses to participate in the microfinance program does not solely depend on whether her friends have chosen to participate. One could assume that participation depends on other properties of each agent, for instance whether she needs a loan, whether she has potential for using such a loan etc. Let us collect all such reasons into one feature variable V_O representing whether an agents is open/responsive to a loan (assuming that $R_R = \{y, n\}$ as well). Another precondition for choosing to participate in the micro-loan program is of course that the agent is actually informed about it. Thus, the diffusion of participation might be modeled by a dynamic transformation $\mathcal{D}_P = (\Phi_P, \mathsf{post}_P)$, where

$$\Phi_P = \{V_I = y \wedge V_O = y\}$$
$$\mathsf{post}_P\big(V_I = y \wedge V_O = y\big)(V_P) = y.$$

With such dynamics we can, for instance, show that if agent j is friend with i and i participates in the program, then after one step of the dynamics $\mathcal{D}_I$, j will be informed, in other words:

$$(@_j\langle F\rangle i \wedge @_i V_P = y) \rightarrow [\mathcal{D}_I]@_j V_I = y.^{15}$$

One can show that after an additional step of the dynamics $\mathcal{D}_P$, j will participate in the program as well. One can also prove more complex properties such as if there is a path of length three from i to j where all agents on the path (including i and j), are open to participation in the program and if i is initially informed, then after three steps of the dynamics, j is participating in the program.

According to [35], participants in the microfinance program were seven times more likely to pass on information about the program than non-participants, while non-participants counted for a third of the passing on of information about the program. This suggests that individuals were informed by non-participants at a much higher rate than they would be if they needed all of their friends to be informed first. It might be natural to assume that an agent gets informed when more than a third of her friends are informed. This kind of preconditions based on thresholds are quite common in the models of network science. Our current logic cannot capture this. However, in the next section, we will briefly discuss extensions of our framework to capture such thresholds preconditions and other interesting traits.

5.4.3 Extension to Numerical Thresholds

In this chapter, we have illustrated how a simple modal logic can be used to reason about a large class of dynamic processes. While our logic has limitations, there are several possible extensions that can make it applicable to larger classes of models from social networks analysis. In this subsection, we briefly sketch how

Axiom:	
$\left(\bigwedge_{s=1}^{n}(@_i\langle F\rangle j_s \wedge @_{j_s}\varphi) \wedge \bigwedge_{1\leq s<t\leq n} @_{j_s}\neg j_t\right) \to @_i\langle{\geq}nF\rangle\varphi$ [1]	n-Bridge
Rule:	
From $\left(\bigwedge_{s=1}^{n}(@_i\langle F\rangle j_s \wedge @_{j_s}\varphi) \wedge \bigwedge_{1\leq s<t\leq n} @_{j_s}\neg j_t\right) \to \psi$, where $i \neq j_s$ and j_s does not occur in φ or ψ (for all s), infer $@_i\langle{\geq}nF\rangle\varphi \to \psi$ [1]	n-Paste
[1] Here $j_1, ..., j_n$ denote distinct nominals.	

Figure 5.4: The additional axioms and rules for the modalities $\langle{\geq}nF\rangle$.

to extend our logic with "numerical threshold modalities". The next subsection briefly discusses other possible extensions which we leave for future research.

In [145, 120, 166] the changes considered depend exclusively on whether "all" or "some" of an agent's neighbors believe/prefer/know something. Thus, changes can be modeled based on the thresholds "at least one" or "all" of an agent's friends satisfying something. Similarly, in what precedes, we have restricted our threshold dynamics to those which are definable using our language $\mathcal{L}_{\mathcal{ND}}$. However, one can argue that many diffusion phenomena involve numerical thresholds. Consider an example where agents only adopt a new technology if at least 5 of their friends/network-neighbors have adopted it already. To capture this idea, we add *numerical threshold modalities*[16] $\langle\geq nF\rangle$ for any $n \in \mathbb{N}$, with the interpretation "at least n of my network-neighbors...". Formally, the semantics is given by:

$$\mathcal{M}, a \models \langle{\geq}nF\rangle\varphi \quad \text{iff} \quad |\{b \in A \mid b \asymp a \text{ and } \mathcal{M}, b \models \varphi\}| \geq n \ ,$$

where $|B|$ denotes the cardinality of the set B.

Axiomatizing this extension of our logic turns out to be surprisingly straightforward due to our use of hybrid logic. We simply add the following modified versions of the Bridge axiom and the Paste rule shown in Figure 5.4. With these, straightforward extensions of the Lindenbaum Lemma and the Truth Lemma, used in the completeness proof, yield completeness of our static logic N extended with the numerical threshold modalities. For our dynamic logic ND adding the numerical threshold modalities are now straightforward as well since the following reduction axiom is easily seen to be valid:

$$[\mathcal{D}]\langle{\geq}nF\rangle\varphi \leftrightarrow \langle{\geq}nF\rangle[\mathcal{D}]\varphi \ .$$

[16] What we call "numerical threshold modalities" here are known as "graded modalities" in the modal logic literature and date back to Kit Fine's paper [78]. They are also known as "qualified number restrictions" in the Description Logic literature [15].

5.5 Conclusion

5.5.1 Summary

In this chapter, we have abstracted away from particular social phenomena in order to design a general modal framework to reason about diffusion in social networks in general:

- We have defined a hybrid logic framework for complex diffusion phenomena,
- We have shown how this new logic allows us to capture any dynamic transformations, that is, any locally determined redistribution of features in social network structures definable in terms of preconditions of our language.
- We have shown how our logic can accommodate several refinements by "plugging-in" some dynamic transformations depending on what type of social phenomena we are focusing on. In particular, we have shown how to apply the setting to the diffusion of micro-loans in villages and to the phenomenon of pluralistic ignorance which was studied in the previous chapter.
- We have given a complete axiomatization of the logic.
- We have presented a simple extension to numerical threshold modalities.

5.5.2 Further Research

Extensions of the logic In addition to the numerical threshold modalities just discussed, one can also consider "proportional threshold modalities". For instance, in the microfinance example of the previous section, it might be more natural to specify that an agent gets informed if one third of her friends are informed. Another example is the dynamics induced by coordination (or anti-coordination) games played in social networks where the threshold to consider will depend on the payoffs involved in the corresponding game [71, Ch. 19].

Formally, we could also add proportional modalities of the form $\langle \geq \frac{p}{q} F \rangle$ for $p, q \in \mathbb{N}$ with $p \leq q$, with the following semantics:

$$\mathcal{M}, a \models \langle \geq \tfrac{p}{q} F \rangle \varphi \quad \text{iff} \quad \frac{|\{b \in A \mid b \asymp a \text{ and } \mathcal{M}, b \models \varphi\}|}{|\{b \in A \mid b \asymp a\}|} \geq \frac{p}{q} \ .$$

In addition to the already mentioned example of microfinance, the extended logic could be used to reason about several standard network analysis issues. For instance, the relationships between the density of clusters of a network structure and the possibility of a complete diffusion or "cascades" under a given threshold (see e.g. Chapter 19 of [71] for a presentation of a theorem without the use of logic and [19] for the same result using logical tools considerably different from our logic). In contrast to the numerical threshold modalities previously discussed, adding the proportional threshold modalities $\langle\geq\frac{p}{q}F\rangle$ to our static logic N requires more work with respect to the axiomatization. Moreover, the given semantics of $\langle\geq\frac{p}{q}F\rangle$ might not be the obvious choice in networks where some agents have infinitely many friends. However, for the case of $\frac{p}{q}=\frac{1}{2}$ a nice solution can be found in [131]. Note that, with respect to the dynamic extension (in the finite case) we still have a straightforward reductions axiom in form of the following validity:

$$[\mathcal{D}]\langle\geq\tfrac{p}{q}F\rangle\varphi \leftrightarrow \langle\geq\tfrac{p}{q}F\rangle[\mathcal{D}]\varphi\ .$$

Another possible extension of our logic, in line with [166], is to add the transitive closure operator F^* of the modality F, with the following semantics:

$$\mathcal{M},a\models F^*\varphi \quad \text{iff} \quad \text{for all } b\in A; a\asymp^* b \text{ implies } \mathcal{M},b\models\varphi,$$

where $\asymp^*$ is the transitive closure of the relation $\asymp$.[17] This "community modality" quantifies over what [166] names an agent's "community", that is, the agent's friends, the agent's friends' friends, the agent's friends' friends' friends, etc. Such a modality, as mentioned earlier in Section 5.2, allows us to express that a network is strongly connected [38, 39, 80]. We have left out this modality in our current logic because it is not essential to the main ideas we wish to convey.

Occasionally, what is of interest is the limit behavior of diffusion processes within social networks. To capture this, a second "transitive closure" modality that we could add to our framework is the transitive closure of the dynamic transformation $\langle\mathcal{D}\rangle$, with the following semantics:

$$\mathcal{M},a\models\langle\mathcal{D}^*\rangle\varphi \quad \text{iff} \quad \text{there is a } k\in\mathbb{N}_0 \text{ such that } \mathcal{M}^{k\mathcal{D}},a\models\varphi.$$

In Section 5.2, we discussed how to describe stability, but our language as such cannot capture stabilization. On the other hand, with the $\langle\mathcal{D}^*\rangle$ modality we can easily express that a network model $\mathcal{M}$ stabilizes under the dynamic transformation $\mathcal{D}$ by the following formula

$$\langle\mathcal{D}^*\rangle\varphi_{stable}(\mathcal{D}).$$

However, sometimes, limiting behavior can be reduced to other properties of the network structures. For instance, [145] gives a characterization of stable and stabilizing models for the particular transformation under consideration, while [62] reduces stabilization of some type of network models to the existence of an odd cycle in the underlying network structure.[18] Again, we have chosen not to

[17]Such transitive closure modalities are also fairly standard in Propositional Dynamic Logic (PDL) [95].

[18]Talking about limiting behavior of social network dynamics using transitive closure modalities is also done in [94] for a particular model of opinion dynamics in social networks. However, the framework of [94] differs considerably from the present framework as it is based on a Fuzzy Logic.

include $\langle \mathcal{D}^* \rangle$ in our logic as it is not essential to the main ideas of this chapter. Moreover, adding $\langle \mathcal{D}^* \rangle$ is again likely to complicate the axiomatization of our logic.

Towards smarter agents Finally, since this chapter has designed a general framework to reason about diffusion dynamics in social networks, a natural next step for further research is to generalize our framework in order to take into account information, and investigate the logical interaction between information and diffusion.

The next two chapters will take this turn. We will propose frameworks incorporating the tools of epistemic logic, to capture how knowledge of the agents interacts with the dynamics of diffusion phenomena. By doing so, we will make a conceptual jump from agents who were so far considered as reacting automatically to their environment in a "bacteria-like" behavior to agents who can reason about each other's behavior, infer each other's private opinions (in Chapter 6), and even anticipate each other's behavior (in Chapter 7).

Chapter 6

Diffusion as Knowledge Source: Hybrid Epistemic Account

In the reminder of this part of the thesis, we add a new component to our picture of social influence and diffusion dynamics: *knowledge*. We will use two different settings to model what agents know about each other and about the social network structure and to investigate how this knowledge interacts with diffusion phenomena. Roughly put, this chapter shows how the dynamics of the diffusion phenomena affects the agents' knowledge; while the next chapter shows how the agents' knowledge affects the diffusion dynamics.

The work presented in the two previous chapters was obtained through sequential generalizations. We added "layers" to the modeling of agents, and extended the corresponding logics accordingly: Chapter 4 designed a dynamic "two-layer" generalization of the hybrid "one-layer" opinion change setting from [120]; and Chapter 5 brought us from "two-layer" opinion change to "many-layers" diffusion phenomena.

By doing so, we have moved not only towards a logic capturing a wider class of network phenomena but also towards more complex agents. This chapter, based on work from [64], goes further in that direction: we now propose an epistemic extension of the "many-layers" hybrid framework for diffusion developed in Chapter 5. This will allow us to model agents who can not only observe each other's behavior and react to it, but who can also reason about the evolution of each other's behavior and "see through it" in this chapter, and even anticipate each other's behavior change (in the next chapter) .

6.1 Introduction

This chapter models what agents in a social network know about each other (and about the network structure), in order to show how diffusion phenomena affect this knowledge, which in turn affects the diffusion dynamics. In particular, we will show how agents can come to gain knowledge about their network neighbors by observing the evolution of their behavior under social influence.

To capture the interaction between knowledge and diffusion phenomena, we will introduce a logic for social networks, knowledge, diffusion, and "learning". Formally, the framework extends the general hybrid logic for diffusion phenomena developed in the previous chapter, enriching it with tools from dynamic epistemic logic: a set of possible situations, some of which are indistinguishable to some of the agents, and some information events. The result is a general two-dimensional dynamic hybrid epistemic setting, containing two types of model transformations to represent the effects social influence: one type to capture ontic changes, the second one to capture epistemic changes. We will illustrate how this new setting allows us to capture the examples we have considered so far.

6.1.1 Outline

The first part of the chapter gives some motivation and examples. The reminder of this section argues in favor of an epistemic treatment of the diffusion phenomena we have been modeling so far, such as the diffusion of opinions under social conformity pressure modeled in Chapter 4.

Section 6.1.2 first recalls our proposal of a "two-layer" opinion change dynamics under social influence from Chapter 4, and presents the main claim of this chapter: our "opaque" two-layered agents might be more "see-through" than we intended. Once we model explicitly the epistemic state of the agents, we can show how agents can sometimes come to learn each other's private opinion, despite the fact that they can only observe their expressed opinion. Section 6.2 illustrates this social influence induced learning process by modeling a simple example, and proposes a possible way to adapt our notion of social influence to a knowledge-dependent one.

The second part of the chapter formalizes the issues discussed so far. Section 6.3 introduces the Logic of Networks, Knowledge, Diffusion, and Learning (NKDL) and Section 6.3.2 illustrates how a fragment of our logic capture all the cases of social influence and learning described so far, both with the knowledge independent and with the new knowledge-dependent social influence (fragment $\mathsf{NKI^RO}$).[1] Section 6.4 gives some conclusions and directions for further research.

6.1.2 "See-through" Agents After All?

Before illustrating how social influence may affect what agents know about each other, we first need a brief "flashback" to some work presented earlier in the thesis.

[1] Our [64] also provides a sound and complete tableau system for a fragment of the new logic, which also contains the first tableau system for the "Facebook Logic" from [145].

Recall the "two-layer" case of opinion dynamics under social influence introduced in Chapter 4. There, we have argued that modeling social influence phenomena requires to make agents "opaque" to each other, in the following sense: typically, while agents can observe each other's public behavior, they cannot observe each other's mental state.

We have then argued that it is this lack of transparency between agents which explains how rational individuals may lead each other to a situation of collective failure, in cases where agents communicate enough to influence each other, but not enough to share all reasons of their behavior with each other. The two social phenomena modeled in Part II of the thesis illustrated this point: Once an informational cascade has started (see Chapter 3), each agent knows the choices made by others preceding her in the sequence, but does not know the justification for these choices.[2] Similarly, in the case of pluralistic ignorance, each agent observes the behavior of others around her, but does not know the reasons behind their behavior.

The possibility of this "opacity" is an essential component of social life. We have then proposed a "two-layer" model of social influence dynamics, where what an agent publicly expresses depends on two factors: what she privately believes herself and what others around her in the network publicly express. According to our assumptions, each agents can observe the behavior of her network neighbors and see this behavior evolve, without having a clue about their private opinions.

This way of conceiving social influence relies on making three implicit epistemic assumptions about our agents who are network-neighbors: 1) they know their own private opinions, 2) they do not know each other's private opinions, and 3) they do know each other's expressed opinions. Using the epistemic logic tools, we will discuss a puzzling consequence of these very assumptions: assuming that the rules of social influence are common knowledge, agents influencing each other may come to "infer" each other's private opinions by simply witnessing how their expressed opinions evolve. Hence, the notion of social influence which we had designed specifically to represent agents who are "opaque" to each other (see Chapter 4.2) may actually allow them to "see through" each other after all. The next section illustrates this effect in a simple example.

6.2 Modeling Knowledge and Social Influence

This section models an example of how agents can learn about other agents' private opinions by reasoning about their public behavior, the rules of social influence, and the involved social network structure.

[2]In our "urn" example, the second agent could infer what justified the choice of the first agent, and the third agent could infer the choice of the second agents (according to our tie-breaking rule), but all later agents were not in a position to know what color of ball the others had observed. They could only observe what color proportion they decided to bid on.

6.2.1 A Two-agent Example

In Chapter 4, we focused on the phenomenon of pluralistic ignorance, as a paradigmatic example of a social phenomenon where the discrepancy between private opinions and public behavior plays an essential role and where learning about others' private opinions might greatly affect the dynamics of the phenomenon. It has been attested that situations of pluralistic ignorance can occur for instance in board meetings [163] and corporate organizations [93]. Given the managerial hierarchy of companies, it is obvious that social influence affects people's behavior and forces them to act strategically. Thus, the corporate setting seems to be a highly relevant and realistic setting for an example:

Consider a small IT startup facing a buy up from a large well-established company. The CEO (Bob) and the CTO (Alice) are deliberating on whether to accept the offer or not. Alice would like to accept the offer as she knows the poor quality of the startup's software. Bob would like to accept the offer as he would like to move on and found another company. Thus, both are of the opinion that they should accept the offer (what we call *pro opinion*). However, Alice believes that Bob sees the company as his life project and is afraid of telling him the true state of their software. Hence, Alice publicly expresses an opinion against accepting the offer (a *contra opinion*). Bob, on the other hand, is a bit more cautious and shows a strong trust in the CTO's ability to judge the true quality of their product. Therefore, he decides to voice a *neutral opinion* towards the buy up. Now, in a world of social influence and uncertainty of others' true motives, what will happen after the initial voicing of opinions by Alice and Bob? How will they individually update their own knowledge about each other's opinion of the other? Will any of them start voicing another opinion than the one they have initially voiced? This is precisely the type of questions which we want our formal model to capture.

6.2.2 Opinion and Uncertainty

In the above example, the set of agents A just consists of Alice (a) and Bob (b), and the issue towards which the agents have opinions is whether to accept the buy up or not. Following the two-layer proposal from [62], we assume that both hold a private "inner" opinion about the buy-out that is either *pro* (we note ip), *contra* (ic), or *neutral* (in). Moreover, they voice their public "expressed" opinion towards the issue that can also be either pro (ep), contra (ec), or neutral (en), In our current example, Alice (a) has an inner pro opinion (ip) and expresses a contra opinion (ec), while Bob has an inner pro opinion (ip) and expresses a neutral opinion (en). We will refer to such a state of affair simply as a *state*. Hence, mathematically, a state s is just a function $s : A \to \{ip, in, ic\} \times \{ep, en, ec\}$. The state of the world representing our example is shown in Figure 6.1.

a:ip,ec
b:ip,en

Figure 6.1: A state where Alice (a) has an inner pro opinion (ip) and publicly expresses a contra opinion (ep), while Bob (b) has an inner pro opinion (ip) and publicly expresses a neutral opinion (en).

We now add to the modeling a representation of the agents' uncertainty, in classical epistemic logic. That is, different possible states of affairs are considered, which each agent may be able to distinguish or not. For instance, in our example we assume that Alice and Bob do not know each other's private opinion. In this case, Alice cannot distinguish between the state shown in Figure 6.1 and a similar state differing only by Bob's inner opinion being neutral or contra. Formally, assume a finite set of possible states W, representing the possible assignments of one inner opinion and one expressed opinion to each agent. For each agent $a \in A$ we will assume an equivalence relation $\sim_a$ on W representing the uncertainty of agent a.[3] Thus, that two states w and v are related by $\sim_a$ means that, for all agent a knows, she cannot distinguish between them. Mathematically, what we currently call a "model", can be viewed as a tuple $\langle A, W, V, (\sim_a)_{a \in A} \rangle$, where $V : W \times A \to \{ip, in, ic\} \times \{ep, en, ec\}$ specifies for each agent and each state, what the agents inner and expressed opinions are. A full fetched epistemic logic to reason about this notion of uncertainty is introduced in Section 6.3.

We make the following simplifying assumptions. Agents know their own inner and expressed opinions, Agents also know their network-neighbors expressed opinions. In our example, Bob and Alice know each other's expressed opinion. Assuming further that Alice and Bob have no information about each other's inner opinions, the resulting epistemic model is as given in Figure 6.2, where the thick box represents the actual state of affairs. In this model, Alice (a) cannot distinguish between two states if she has the same inner and expressed opinions in both and Bob's expressed opinions also coincide (his inner opinions might differ though). This indistinguishability is represented in the figure as a dashed line between the states marked by a. Since the uncertainty relation is an equivalence relation it is also reflexive and transitive, but to ease the reading of figures we will not draw the corresponding reflexive and transitive dashed lines.

What do agents know about the network structure? It seems natural to allow for uncertainty with respect to social relations, as agents might not know who are their friends' friends, for instance. To allow for uncertainty about the network structure, each possible state will come equipped with its own network: we have an irreflexive and symmetric social relation $\asymp_w$ on A for each state $w \in W$. Then, $a \asymp_w b$ means that agent a and agent b are socially related in the state w. We

[3]We therefore use the standard epistemic logic definition of knowledge as $S5$ modality, system of epistemic logic.

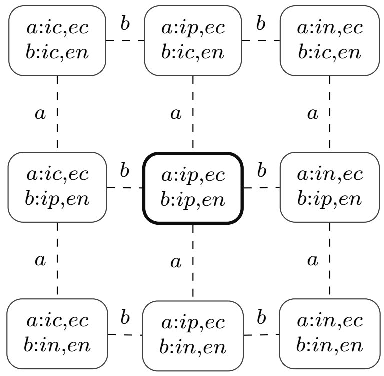


Figure 6.2: In this figure, nodes represent possible states of affairs. The actual state is highlighted by a thicker frame. A dashed line with superscript a (resp. b) represents the indistinguishability between states for agent a (resp. b). We omit reflexive and transitive edges between nodes.

make only one assumption about the interaction between the relations $\sim_a$ and $\asymp_w$, namely that all agents know to whom they are socially related to. Formally, this means that for all states $w, v \in W$ and all agents $a, b \in A$, if $w \sim_a v$, then $a \asymp_w b$ if, and only if, $a \asymp_v b$.

In our running example, the network structure is assumed to be trivial: Alice and Bob are the only two agents and they are connected to each other. For this reason, we have also omitted the representation of the network structure in the figures 6.1 and 6.2. However, we will consider an example with a non-trivial network structure later on.

6.2.3 Opinion and Uncertainty Change

The models so far only represent a static picture of the world. We will add two substantially different types of dynamics: social influence dynamics and learning dynamics. The social influence dynamics will specify how agents change their expressed opinion, as discussed in Chapter 4. The learning dynamics will specify how to update the epistemic dimension of our models, that is, the agents' indistinguishability relations $\sim_a$.

For the social influence dynamics, we will use the notion of "two-layer" social influence given by Def. 4.2.1, extensively discussed in Chapter 4. This notion is independent from the epistemic structure of the model. Let us recall here two important simplifying assumptions: 1) only the agents's expressed opinions are affected by social influence (i.e, their inner opinion does not change), and 2) our agents tend to express sincerely their inner opinion, whenever the social pressure does not prevent them to do so. To avoid confusing the reader by the different terms used in different chapters, we reformulate Def. 4.2.1 in terms of "pro", "contra", and neutral opinions (instead of "inner belief that φ","inner belief that

$\neg\varphi$" and "inner undecidedness"):

6.2.1. Definition. [2-layer social influence (reformulated from Def. 4.2.1)] The following rules constitute the notion of *two-layer social influence*

- Strong Influence: If all the friends of an agent express a pro (contra) opinion, the agent will fall in line and express a pro (contra) opinion in the next round.

- An agent is brave enough to sincerely express her inner pro (contra) opinion if she has some support from her friends (i.e. at least one friend who also expresses a pro (contra) opinion) or she has no opposing friends expressing the opposite contra (pro) opinion (i.e all friends express a pro (contra) or neutral opinion).

- If an agent has an inner neutral opinion and is not under strong influence to express either a pro or contra opinion, she will express a neutral opinion.

- (Revised) weak influence: If an agent has an inner pro (contra) opinion, at least one opposing friend, and no support (i.e. at least one friend expresses a contra (pro) opinion and the other friends either express a contra (pro) or neutral opinion), she will express a neutral opinion.

Let us illustrate this notion by returning to our buy-up example. As this notion of social influence is independent of the epistemic structure of our model we will ignore the epistemic structure for now and recall the state of Figure 6.1. Note that Alice has no opposing friend and thus, according to the above-mentioned rules of social influence, she will sincerely express her pro opinion in the next round. Bob, on the other hand, is under strong influence to express a contra opinion as all his friends (Alice) express a contra opinion. Hence, the initial situation of step 0 will evolve into step 1 of Figure 6.3. Now Alice is under strong influence to express a contra opinion and Bob is under strong influence to express a pro opinion. Thus, we obtain Step 2 of Figure 6.3, where Alice is under strong influence to express a pro opinion and Bob is under strong influence to express a contra opinion. Hence, we will move on to Step 3, at which point Alice and Bob have swiched expressed opinions.

$$\begin{array}{ccccccccc}
\text{Step 0:} & & \text{Step 1:} & & \text{Step 2:} & & \text{Step 3:} & & \\
\boxed{\begin{matrix} a{:}ip,ec \\ b{:}ip,en \end{matrix}} & \rightsquigarrow & \boxed{\begin{matrix} a{:}ip,ep \\ b{:}ip,ec \end{matrix}} & \rightsquigarrow & \boxed{\begin{matrix} a{:}ip,ec \\ b{:}ip,ep \end{matrix}} & \rightsquigarrow & \boxed{\begin{matrix} a{:}ip,ep \\ b{:}ip,ec \end{matrix}} & \rightsquigarrow & \cdots
\end{array}$$

Figure 6.3: The evolution of opinions under social influence ($\rightsquigarrow$) in the running example.

This oscillation will continue: Alice and Bob will keep switching between expressing a pro and a contra opinion towards the buy up. However, this may seem counterintuitive and brings us to the key claim of our chapter: *Agents can learn about each others' private opinions by observing each others behavior (expressed opinion) and adapt their own behavior accordingly, essentially resulting in an alternative, more enlightened, social influence dynamics.* Before introducing the alternative notion of social influence, let us formally introduce the learning that can happen in the above example.

Recall the initial situation of our example with the uncertainty explicitly present, as illustrated by Figure 6.2. As the rules of social influence are independent from the epistemic structure, we can apply the rules of social influence to all the states in Figure 6.2. This will result in the updated model shown Step 1 of Figure 6.4. We could continue to apply the rules of social influence again and again, however, for the state we really care about, the actual state in the middle, the resulting dynamics will just be the one shown in Figure 6.3.

Alternatively, we can take a closer look at the epistemic structure and contemplate on whether the social influence update should also induce a change in the epistemic structure. This seems highly plausible indeed, if we assume that Alice and Bob are aware of the rules of social influence and are always aware of each other's expressed opinion.

Consider the situation from the viewpoint of Bob in the actual state in Step 1. Here Bob still does not know Alice's inner state. Nevertheless, after seeing that Alice now expresses a pro opinion, Bob can actually learn something, if he assumes that Alice obeys the rules of social influence. If Alice had had an inner contra opinion, she would now express such a contra opinion. Similarly, if she had had an inner neutral opinion, she would now express a neutral opinion. In other words, by counterfactual reasoning, Bob can now exclude the middle left and the middle right state in Step 1 of Figure 6.4. Note that this potential counterfactual reasoning by Bob is explicitly represented in our models by the fact that Alice has a different expressed opinion in each of the middle states and by the assumption that Bob can see Alice's expressed opinion. Thus, while Bob could not initially distinguish between any of the states in the middle row, he now can. Note that, this argument also works if Bob's inner opinion had been different (that is if the actual state had been the upper middle or the lower middle state). In conclusion, Bob's learning should correspond to the removal of the horizontal indistinguishability relations labeled b in Step 1 of Figure 6.4, since Alice has different expressed opinions in each of them. Formally, this learning operation simply corresponds to letting agents distinguish between states where their friends have different expressed opinions.

By a closer inspection, it becomes clear that Alice cannot perform a similar learning. Hence, the possible learning of the agents will result in the model of Step 1' of Figure 6.4. That is, Bob has learned that Alice has an inner opinion in favor of the buy up, while Alice did not learn anything about Bob's private

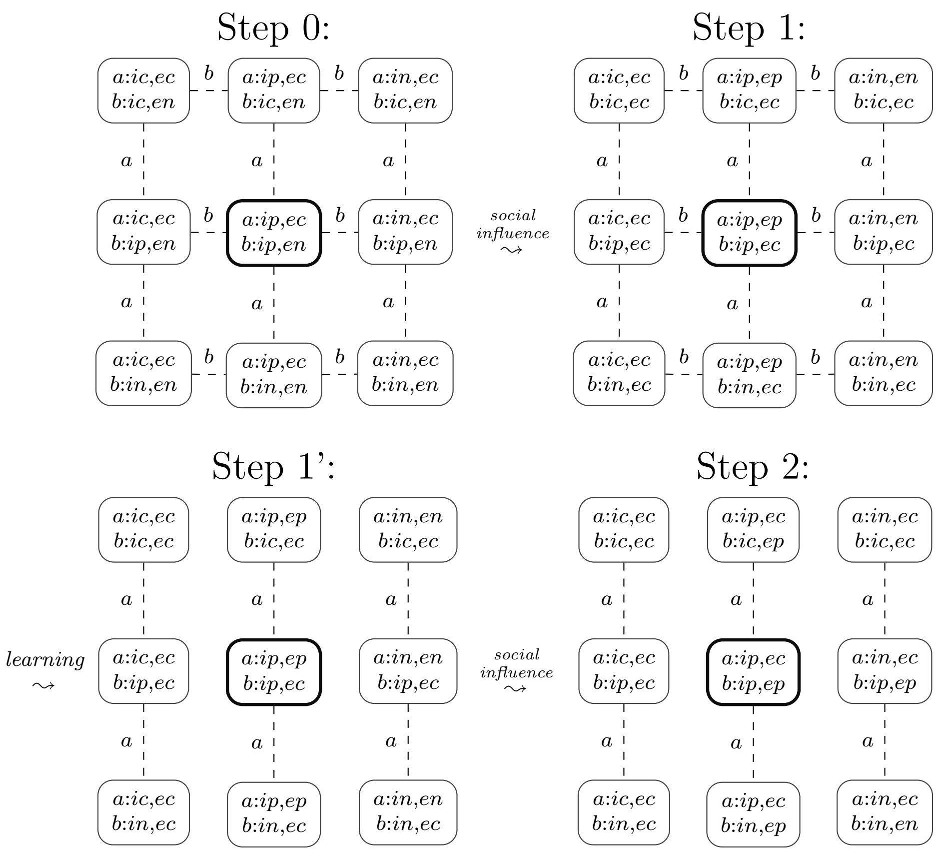


Figure 6.4: The development of the example of Figure 6.2 under alternating applications of the rules of social influence and learning.

opinion towards the buy up. Now, applying another round of the rules of social influence would result in a model that looks like Step 2 of Figure 6.4 as the rules are independent of the epistemic structure. Thus, while Bob has learned Alice's private opinion, the resulting opinion dynamics is still the same oscillating one (as initially pictured in Figure 6.2).

6.2.4 Uncertainty-dependent Opinion Change

In the previous sections, we have shown how agents can learn about each other's private inner opinion by observing the evolution of their expressed opinion, assuming that all agents obey our two-layer influence rules given by Def. 4.2.1.

One could now argue that the two layer notion of social influence, relying on the very idea that agents do *not* have access to each other's private opinions, and according to which agents are influenced *only* by the expressed opinions of their neighbors, does not make much sense anymore when agents might know each other's private opinions. When agents know the private opinions of others, should this additional information not affect how they react to social conformity pressure? For instance, in a situation where all my neighbors express a contra opinion, would the pressure to express a contra opinion not be considerably weakened if I knew that none of them actually holds a contra opinion privately? This section proposes an alternative notion of social influence which takes into account the private opinion of neighbors, *whenever it is known*.

Consider again Step 1' of Figure 6.4. Now that Bob knows that Alice's inner opinion is pro, he knows that he has private support for the buy up and it seems only natural that he should thus express sincerely his inner pro opinion publicly for all future instances. In Step 2 of Figure 6.4 Bob does indeed express a pro opinion. However, under the previously defined rules of social influence, we will see an infinitely oscillating pattern of expressing pro and contra opinion by Bob, as discussed following Figure 6.2. What we claim here is that, given what he has learned, Bob should not show this pattern of infinite oscillation, but instead stick to expressing a pro opinion from now on.

It seems counterintuitive that an agent would keep feeling the same level of conformity pressure if he knows (as Bob in our example) that others around him privately agree with his opinion. To compensate for this counterintuitive consequence of our two-layer definition of social influence, we propose to adapt it to take into account what agents know about each other's private opinion. We will refer to this new notion of social influence as "*reflective social influence*":

6.2.2. Definition. [Reflective social influence] The following rules constitute the notion of *reflective social influence*:

- If an agent has an inner pro (contra) opinion and knows that she has some private support (she knows that she has a friend with an inner pro (contra) opinion as well) or she knows that she has no privately opposing friends (she knows that none of her friends has a inner contra (pro) opinion), then she will express sincerely her inner pro (contra) opinion.

- If an agent has an inner neutral opinion and she is not truly alone with this opinion (she knows she has a friend who also has an inner neutral opinion), then she will express a neutral opinion.

- If none of the above cases apply, then the old rules of social influence (Definition 6.2.1) apply.

Intuitively, this notion of social influence simply makes sure that when an agent knows that she is not truly (not sincerely) under strong or weak influence to express anything else, she will express sincerely her inner opinion. In all other cases, our previous two-layered definition still applied.

We have now seen how agents can sometimes learn each other's private opinion by observing their behavior and we have argued for a new notion of social influence taking into account this observation.

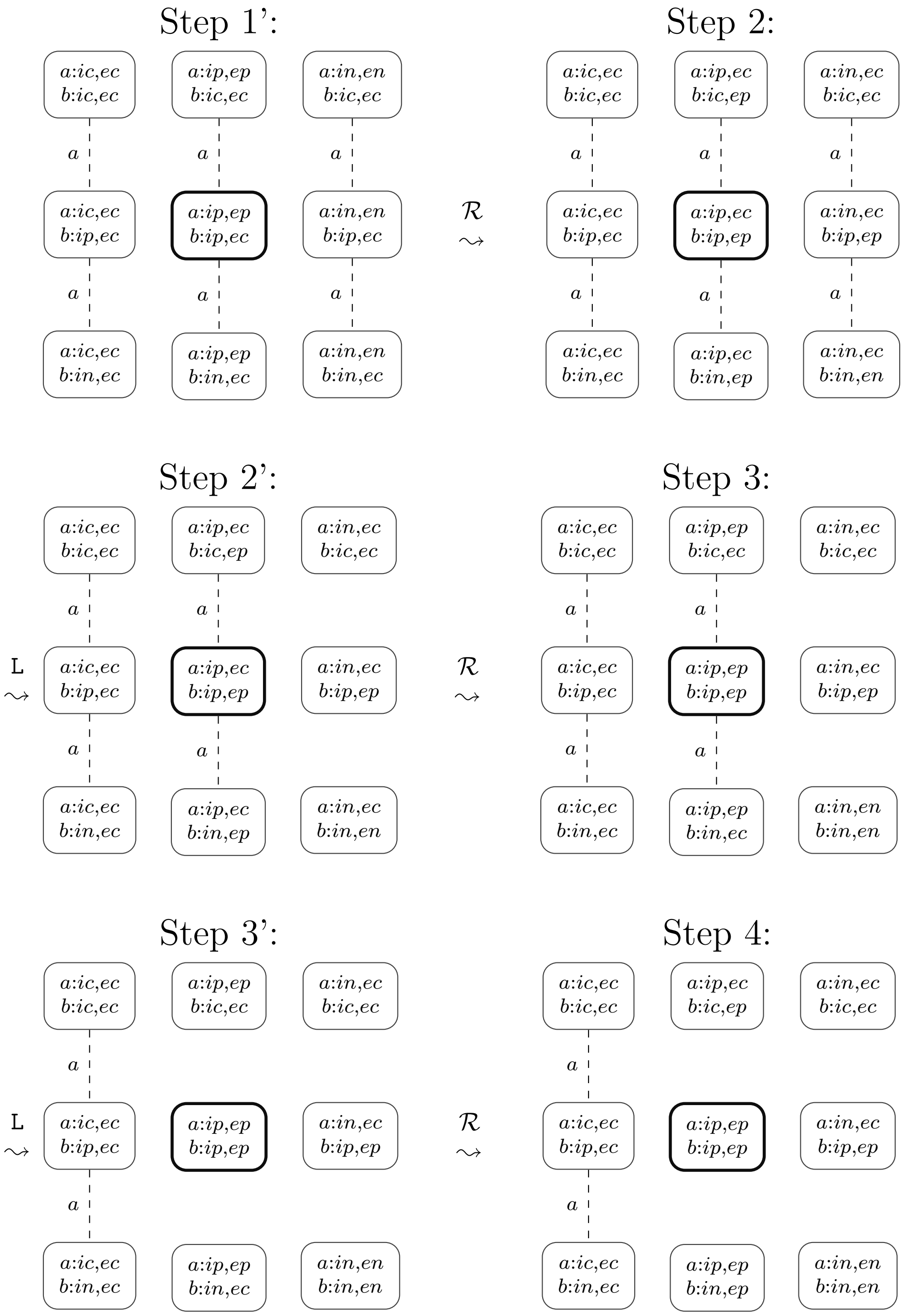


Figure 6.5: The buy-up example fro Fig. 6.3 with the new social influence operator $\mathcal{R}$ defined in Def. 6.2.2. Step 0, Step 1, Step 1' and Step 2 are the same as with the initial influence operator $\mathcal{I}$ (see Fig. 6.4). However, in Step 2, since b knows that a's private opinion is ip, b will continue expressing sincerely his pro opinion in Step 3 instead of aligning on a's expressed opinion. Note that the situation in the actual world stabilizes (unlike in Fig. 6.4), but not the situation in the world just above it, in which agents will keep oscillating, because they privately disagree.

To conclude this section, let us briefly illustrate why the social network structure in our models matters.

Consider again our example. It might be the case that Alice does not entirely form her publicly expressed opinion based solely on her own true opinion and the expressed opinion of Bob, but she also takes the public opinion of Carol, the startup's lead developer, into account. Moreover, Bob may very well be aware of this. Thus, the social network structure of "potentially direct influence of opinions" is as follows:

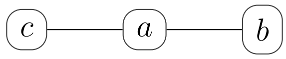


Figure 6.6: The network of Alice, Bob and Carol.

That is, Alice influences and is influenced by both Bob and Carol. The crucial modeling assumption is that Bob and Carol cannot observe each other's expressed opinion (and vice versa). Assume now that Carol has an inner contra opinion and expresses a contra opinion.

Now, as there is no knowledge of inner opinions, the first step of social influence will result in Bob expressing a contra opinion (as before), Carol expressing a contra opinion as well, and Alice expressing a neutral opinion (she is now under weak influence). By the new expressed behavior of Alice, Bob cannot learn anything anymore, since he does not know what opinion Carol expressed. In fact, for all Bob knows, all possible inner opinions for Alice are consistent with her now expressing a neutral opinion: If Alice had an inner neutral opinion and Carol expressed a neutral opinion, Alice would also express a neutral opinion. Had Alice had a inner contra opinion and Carol had expressed a pro opinion, Alice would also express a neutral opinion. Thus, without knowing anything about Carol's expressed opinion, Bob simply cannot learn anything about Alice's inner opinion at this point. This shows the importance of the particular social network structure for what agents can learn.

6.3 Logic of Networks, Knowledge, Diffusion, and Learning

The previous section illustrates how network-neighbors under social influence can sometimes discover each other's private opinions even though they can only observe each other's expressed opinions. In this section, we introduce a general formal framework to reason about such phenomena, named *Logic of Networks, Knowledge, Diffusion, and Learning* (or NKDL).

The logic will incorporate ideas and tools from different frameworks. We extend the dynamic *Logic for Diffusion in Social Networks* ND introduced in the previous chapter, in two ways. First, by adding an *epistemic dimension* to it, with a standard knowledge operator K, obtaining a two-dimensional setting in the spirit of the "Facebook Logic" setting proposed by [145]. Secondly, by adding a *dynamic learning modality*, inspired by the link cutting updates of [49], the graph modifiers of [12], and the Arrow Update Logic of [113]. In other words, our setting is at the same time a *dynamic* extension of the epistemic "Facebook logic" of [145] and an *epistemic* extension of our dynamic "Logic of Social Networks Diffusion" from Chapter 5.

Section 6.3.1 first introduces the general framework (NKDL), and Section 6.3.2 shows how the examples discussed in Section 6.2 can be captured by a fragment ($\mathsf{NKI^RO}$) of this framework.

6.3.1 Introducing the Full Logic (NKDL)

Let us briefly recall the key ingredients from the setting of *Logic for Diffusion in Social Networks* ND developed in the previous chapter. The language combines tools from hybrid logic (nominals i and satisfaction operators $@_i$), Fmodality F quantifying over network-neighbors, a dynamic modality representing diffusion, and a special type of atomic propositions, called *feature propositions*.

The feature propositions (coming from a fixed set FP), given by Definition 5.2.1, are used to describe the agents' state, that is, to express which value is assigned to each of her relevant properties (or variables) under consideration. A feature proposition $V_l = r$ reads as "property V_l takes value r (at me)". In particular, in the cases of opinion distribution considered in the previous sections, the only two relevant properties of an agent are her inner opinion (let us denote it by V_i) and her expressed opinion (V_e). Therefore, the six propositions used so far ip, ic, in, ep, ec, en can be considered as abbreviations for the feature propositions resulting from the assignment of exactly one out of three values to two properties of agents, respectively: $V_i = p, V_i = c, V_i = n, V_e = p, V_e = c, V_e = n$. Hence, for instance, in (or "$V_i = n$") reads "my inner opinion is neutral".

A set DT (to be defined in Def. 6.3.3) contains all the possible ways of reassigning values to properties which are definable in terms of preconditions expressible in the language and dynamic modality $[\mathcal{D}]$ for each dynamic transformation $\mathcal{D}$, such that $[\mathcal{D}]$ reads "after applying the transformation $\mathcal{D}$". In particular, going back again to our example of Section 6.2, the reflexive and non-reflexive influence operators are such possible transformations: they output a unique new opinion state of each agent in the network, depending on the distribution of opinions (and of knowledge) in the input model.

To add the epistemic dimension, we add a standard knowledge modality K to our language. We also add a dynamic modality $[\mathcal{L}]$, for *learning*, for a given finite set $\mathcal{L}$ of formulas of our language (to be defined in Def. 6.3.1). This corresponds to a second type of model transformers (epistemic change), a "learning update", which outputs a model where ontic facts remain unchanged compared to the input model, but where what the agents know might differ. $[\mathcal{L}]$ will thus read "after the learning update $\mathcal{L}$".

Combining the above ingredients, our language $\mathcal{L}_{\mathcal{NKDL}}$ allows for describing diffusion, knowledge, and learning in social networks:

6.3.1. DEFINITION. [Syntax of $\mathcal{L}_{\mathcal{NKDL}}$]

$$\varphi \ ::= \ V_l = r \ | \ i \ | \ \neg\varphi \ | \ \varphi \wedge \varphi \ | \ F\varphi \ | \ @_i\varphi \ | \ [\mathcal{D}]\varphi \ | \ K\varphi \ | \ [\mathcal{L}]\varphi$$

where $V_l = r \in \mathsf{FP}$, $i \in \mathsf{NOM}$, $\mathcal{D} \in \mathsf{DT}$, and $\mathcal{L} \subseteq \mathcal{L}_{\mathcal{NKDL}}$ is finite.[4] [5]

The definition of assignments and full assignments is kept as in Definition 5.2.3 and the two dimensional models are obtained by simply adding the new epistemic components (possible worlds and subjective indistinguishability relations) to Def. 5.2.4:

6.3.2. DEFINITION. [Epistemic Network Model (ENM)] A *model* is a tuple $\mathcal{M} = (A, W, (\asymp_w)_{w \in W}, (\sim_a)_{a \in A}, g, V)$, where:

- A is a non-empty set of agents,
- W is a non-empty set of possible worlds,
- $\asymp_w$ is an irreflexive, symmetric, binary relation on A, for each $w \in W$ (representing the network structure at the world w),
- $\sim_a$ is an equivalence relation on W for each $a \in A$ (representing the uncertainty of a),
- $\asymp_w$ and $\sim_a$ satisfy: for all $w, v \in W$, $a, b \in A$, if $w \sim_a v$, then $a \asymp_w b$ iff $a \asymp_v b$ (the agents know who their friends are).
- $g : \mathsf{NOM} \to A$ is a function assigning an agent to each nominal,
- $V : W \times A \to \mathcal{V}^{full}$ is a valuation assigning a full assignment $\nu(w, a)$ in each possible world $w \in W$, to each agent $a \in A$, i.e. a world-dependent complete specification of the features of each agent in the network.

[4] To avoid circular definitions, the set $\mathcal{L}$ should be forbidden to have precondition formulas involving $[\mathcal{L}]$ itself. Nevertheless, we can allow formulas of $\mathcal{L}_{\mathcal{NKDL}}$ in $\mathcal{L}$ constructed at an "earlier stage" in a simultaneous inductive definition of learning modalities and the language $\mathcal{L}_{\mathcal{NKDL}}$. Similar restrictions will be imposed in the definition of DT below in Def. 6.3.3.

[5] We use the standard abbreviations for $\vee$, $\to$, and $\leftrightarrow$. Moreover we will denote the dual of F by $\langle F\rangle$ and the dual of K by $\langle K\rangle$, in other words $\langle F\rangle\varphi := \neg F\neg\varphi$ and $\langle K\rangle\varphi := \neg K\neg\varphi$.

On the dynamic side, we define now the two model transforming operations, the first one for modeling the redistribution of features of agents (ontic change), in the same way as in the previous chapter, the second one for modeling what agents come to learn (epistemic change).

The set DT of dynamic transformations (ontic changes) are defined as in Def. 5.2.6 but with formulas from our newly extended language:

6.3.3. Definition. [Dynamic transformations (on ENM)] A *dynamic transformation* is a pair $\mathcal{D} = (\Phi, \mathsf{post})$ consisting of a non-empty finite set Φ of pairwise unsatisfiable formulas (from the language $\mathcal{L}_{\mathcal{NKDL}}$ defined in Def. 6.3.1)[6] and a post-condition function $\mathsf{post} : \Phi \to \mathcal{V}$. The set Φ will be referred to as "preconditions". Given a precondition $\varphi \in \Phi$, we will call the assignment $\mathsf{post}(\varphi) \in \mathcal{V}$ the "post-condition" of φ.

The corresponding transformation updates are adapted to the two-dimensional models in the obvious way, by performing the transformation update given by Def. 5.2.8 in each of the possible states:

6.3.4. Definition. [Transformation updates (on ENM)]
Given a model $\mathcal{M} = (A, W, (\asymp_w)_{w\in W}, (\sim_a)_{a\in A}, g, V)$ and a dynamic transformation $\mathcal{D} = (\Phi, \mathsf{post})$, the *updated* model under the transformation $\mathcal{D}$ is $\mathcal{M}^{\mathcal{D}} = (A, W, (\asymp_w)_{w\in W}, (\sim_a)_{a\in A}, g, V')$, where V' is defined by:

$$\nu'(w,a)(l) \;=\; \begin{cases} \mathsf{post}(\varphi)(l) & \textit{if there is a } \varphi \in \Phi \textit{ such that } \mathcal{M}, w, a \models \varphi \\ & \textit{and } l \in \textit{dom}(\mathsf{post}(\varphi)) \\ \nu(w,a)(l) & \textit{otherwise} \end{cases} \tag{6.1}$$

for all all $w \in W$, all $a \in A$, and all $l \in \{1, ..., n\}$.

Let us now turn to the second type of model transformation process, the one representing what agents learn:

6.3.5. Definition. [Learning updates $\mathcal{M}^{\mathcal{L}}$]
Given a model $\mathcal{M} = (A, W, (\asymp_w)_{w\in W}, (\sim_a)_{a\in A}, g, V)$ and a finite set of formulas $\mathcal{L} \subseteq \mathcal{L}_{\mathcal{NKDL}}$ (such that it does not involve $[\mathcal{L}]$ itself), the learning updated model $\mathcal{M}^{\mathcal{L}}$ is $(A, W, (\asymp_w)_{w\in W}, (\sim'_a)_{a\in A}, g, V)$, where $\sim'_a$ is defined by:

$$\begin{array}{lll} w \sim'_a v & \textit{iff} & w \sim_a v \textit{ and, there is no agent } b \textit{ such that } a \asymp_w b \textit{ and;} \\ & & \mathcal{M}, w, b \models \varphi \textit{ and } \mathcal{M}, v, b \models \neg\varphi, \textit{ or} \\ & & \mathcal{M}, w, b \models \neg\varphi \textit{ and } \mathcal{M}, v, b \models \varphi, \textit{ for a } \varphi \in \mathcal{L}. \end{array}$$

[6]Similarly as for modalities $[\mathcal{L}]$ in Def 6.3.1, the dynamic transformation $\mathcal{D} = (\Phi, \mathsf{post})$ should not contain precondition formulas in Φ involving $[\mathcal{D}]$ itself, but can contain formulas of $\mathcal{L}_{\mathcal{NKDL}}$ constructed on an "earlier stage" in a simultaneous inductive definition of dynamic transformations and the language $\mathcal{L}_{\mathcal{NKDL}}$. In other words, one should view Definition 6.3.3 and Definition 6.3.1 as one simultaneous recursive definition.

Informally, this means that we should cut all the indistinguishability links of an agent that relate any pair of states among which the satisfaction of some $\varphi \in \mathcal{L}$ differs at some of her neighbors. In our example above (Fig. 6.4), Bob learns Alice's private opinion because he can distinguish between states in which her expressed opinion is different. This corresponds to saying that $ep, en, ec \in \mathcal{L}$, which is precisely what we assumed as our starting point: agents know the expressed opinions of their network neighbors.[7]

Now that we have defined our models and our two types of model transformers, we give the full semantics of $\mathcal{L}_{\mathcal{NKDL}}$:

6.3.6. Definition. [Semantics of $\mathcal{L}_{\mathcal{NKDL}}$]
Given a model $\mathcal{M} = (A, W, (\asymp_w)_{w\in W}, (\sim_a)_{a\in A}, g, V)$, $a \in A$, $w \in W$, and formulas $V_l = r, i,$ *and* $\varphi \in \mathcal{L}_{\mathcal{NKDL}}$, the truth of φ at (w, a) in $\mathcal{M}$ is defined inductively by:

$$\begin{array}{lll}
\mathcal{M}, w, a \models V_l = r & \textit{iff} & \nu(w,a)(l) = r \\
\mathcal{M}, w, a \models i & \textit{iff} & g(i) = a \\
\mathcal{M}, w, a \models \neg\varphi & \textit{iff} & \mathcal{M}, w, a \not\models \varphi \\
\mathcal{M}, w, a \models \varphi \wedge \psi & \textit{iff} & \mathcal{M}, w, a \models \varphi \textit{ and } \mathcal{M}, w, a \models \psi \\
\mathcal{M}, w, a \models F\varphi & \textit{iff} & \forall b \in A; a \asymp_w b \Rightarrow \mathcal{M}, w, b \models \varphi \\
\mathcal{M}, w, a \models @_i\varphi & \textit{iff} & \mathcal{M}, w, g(i) \models \varphi \\
\mathcal{M}, w, a \models K\varphi & \textit{iff} & \forall v \in W; w \sim_a v \Rightarrow \mathcal{M}, v, a \models \varphi \\
\mathcal{M}, w, a \models [\mathcal{D}]\varphi & \textit{iff} & \mathcal{M}^{\mathcal{D}}, w, a \models \varphi \\
\mathcal{M}, w, a \models [\mathcal{L}]\varphi & \textit{iff} & \mathcal{M}^{\mathcal{L}}, w, a \models \varphi
\end{array}$$

where $\mathcal{M}^{\mathcal{D}}$ is given by Def. 6.3.4 and $\mathcal{M}^{\mathcal{L}}$ by Def. 6.3.5. We say that a formula φ is true in a model $\mathcal{M} = (A, W, (\asymp_w)_{w\in W}, (\sim_a)_{a\in A}, g, V)$ if $\mathcal{M}, w, a \models \varphi$ for all $w \in W$ and all $a \in A$. We denote this by $\mathcal{M} \models \varphi$. We say that a formula φ is valid if $\mathcal{M} \models \varphi$ for all models $\mathcal{M}$ and denote this by $\models \varphi$. The resulting logic will be referred to as the *logic of knowledge, diffusion and learning* and will be denoted NKDL.

How general is our framework? We did impose that agents know who their neighbors are, but not that the network structure is common knowledge (although it is in the examples throughout this chapter). We also restricted our learning events to "learning about neighbors": the learning update will not force agents to come to know the formulas in $\mathcal{L}$, but whether those formulas are satisfied *at their neighbors*. Finally, we required that $\sim_a$ be an equivalence relation. The logic of the knowledge modality K will thus be standard S5. Our framework is therefore general enough to reason both about the diffusion of any number of agents' features dictated by any definable dynamic transformation and about how agents learn features of their neighbors by observing how some of their (possibly other) features change.

[7]Note that our framework does not rely on the assumption that the network structure is common knowledge.

In the next section, we will show how the examples of social influence relying on private versus expressed opinions and learning, as introduced in the previous sections, can be captured by our general framework (NKDL). More precisely, we will define a fragment ($\mathsf{NKI^RO}$) of the framework which is sufficient for modeling such cases.

6.3.2 Capturing our Examples (Fragment $\mathsf{NKI^RO}$)

In this section, we show how the above defined general framework (NKDL) allows for modeling the cases of social influence and learning introduced in Section 6.2. We define the fragment $\mathsf{NKI^RO}$ of our logic: the logic of *Networks, Knowledge, Influence, Reflexion, and Observation*.

To capture our examples, we only need to model the two relevant properties of our agents: their inner opinion V_i and expressed opinions V_e, each of which takes one of the three values p, c, or n. That is, we fix the set of feature propositions of our language to be: $FP = \{V_i = p, V_i = c, V_i = n, V_e = p, V_e = c, V_e = n\}$, or their above-mentioned respective abbreviations: $\{ip, ic, in, ep, ec, en\}$.

We then have to fix the dynamic transformations representing the two notions of social influence we have chosen for opinion changes in our examples: the "two-layer social influence" given in Def. 6.2.1 (and 4.2.1), and described case by case in Figure 4.1; and the knowledge-dependent "reflexive social influence" from Def. 6.2.2, described in a similar table below:

	Inner state	Knowledge state	$\langle F\rangle ep$	$\langle F\rangle ec$	$\langle F\rangle en$	Update
1	ip	$K\langle F\rangle ip \vee K\neg\langle F\rangle ic$	—	—	—	$\rightsquigarrow ep$
2	ic	$K\langle F\rangle ic \vee K\neg\langle F\rangle ip$	—	—	—	$\rightsquigarrow ec$
3	in	$K\langle F\rangle in$	—	—	—	$\rightsquigarrow en$
4	ip	$\neg(K\langle F\rangle ip \vee K\neg\langle F\rangle ic)$	see Fig. 4.1	see Fig. 4.1	see Fig. 4.1	see Fig. 4.1
5	ic	$\neg(K\langle F\rangle ic \vee K\neg\langle F\rangle ip)$	see Fig. 4.1	see Fig. 4.1	see Fig. 4.1	see Fig. 4.1
6	in	$\neg K\langle F\rangle in$	see Fig. 4.1	see Fig. 4.1	see Fig. 4.1	see Fig. 4.1

Figure 6.7: Case by case description of the "reflexive social influence" transformation. "—" means that any value may be inserted here.

We now define the two corresponding dynamic transformations, denoted by $\mathcal{I}$ and $\mathcal{R}$, respectively:

6.3.7. Definition. [Two-layer social influence dynamic transformation $\mathcal{I}$] The dynamic transformation $\mathcal{I}$ is given by $\mathcal{I} = (\Phi_{\mathcal{I}}, \mathsf{post}_{\mathcal{I}})$, such that:

$$\begin{aligned} \Phi_{\mathcal{I}} = \ \{ & (ip \wedge (\langle F\rangle ep \vee [F]en)) \vee [F]ep, \\ & (ic \wedge (\langle F\rangle ec \vee [F]en)) \vee [F]ec, \\ & in \wedge (\neg[F]ep \wedge \neg[F]ec) \ \} \end{aligned}$$

$$\begin{aligned} \mathsf{post}_{\mathcal{I}}\big((ip \wedge (\langle F\rangle ep \vee [F]en)) \vee [F]ep\big)(e) &= p, \\ \mathsf{post}_{\mathcal{I}}\big((ic \wedge (\langle F\rangle ec \vee [F]en)) \vee [F]ec\big)(e) &= c, \\ \mathsf{post}_{\mathcal{I}}\big(in \wedge \neg[F]ep \wedge \neg[F]ec\big)(e) &= n. \end{aligned}$$

6.3.8. Definition. [Reflexive influence dynamic transformation $\mathcal{R}$] The dynamic transformation $\mathcal{R}$ is given by $\mathcal{R} = (\Phi_{\mathcal{R}}, \mathsf{post}_{\mathcal{R}})$, such that:

$$\begin{aligned} \Phi_{\mathcal{R}} = \ \{ & (K\langle F\rangle ip \vee K\neg\langle F\rangle ic) \wedge ip, \\ & (K\langle F\rangle ic \vee K\neg\langle F\rangle ip) \wedge ic, \\ & K\langle F\rangle in \wedge in, \\ & \neg(K\langle F\rangle ip \vee K\neg\langle F\rangle ic) \wedge ((ip \wedge (\langle F\rangle ep \vee [F]en)) \vee [F]ep), \\ & \neg(K\langle F\rangle ic \vee K\neg\langle F\rangle ip) \wedge ((ic \wedge (\langle F\rangle ec \vee [F]en) \vee [F]ec), \\ & \neg K\langle F\rangle in \wedge ((in \wedge \neg[F]ep \wedge \neg[F]ec) \ \} \end{aligned}$$

$$\begin{aligned} \mathsf{post}_{\mathcal{R}}\big((K\langle F\rangle ip \vee K\neg\langle F\rangle ic) \wedge ip\big)(e) &= p, \\ \mathsf{post}_{\mathcal{R}}\big((K\langle F\rangle ic \vee K\neg\langle F\rangle ip) \wedge ic\big)(e) &= c, \\ \mathsf{post}_{\mathcal{R}}\big(K\langle F\rangle in \wedge in\big)(e) &= n, \\ \mathsf{post}_{\mathcal{R}}\big(\neg(K\langle F\rangle ip \vee K\neg\langle F\rangle ic) \wedge ((ip \wedge (\langle F\rangle ep \vee [F]en)) \vee [F]ep)\big)(e) &= p, \\ \mathsf{post}_{\mathcal{R}}\big(\neg(K\langle F\rangle ic \vee K\neg\langle F\rangle ip) \wedge ((ic \wedge (\langle F\rangle ec \vee [F]en) \vee [F]ec)\big)(e) &= c, \\ \mathsf{post}_{\mathcal{R}}\big(\neg K\langle F\rangle in \wedge ((in \wedge \neg[F]ep \wedge \neg[F]ec)\big)(e) &= n. \end{aligned}$$

Those two dynamic transformations represent the two type of opinion change induced by social influence that we had considered in our examples. The set DT is thus fixed to be $\mathsf{DT} = \{\mathcal{I}, \mathcal{R}\}$. We also need to capture formally what agents can observe of each other. Therefore, we fix the learning set to be $\mathcal{O} = \{ep, ec, en\}$, to represent the fact that after the corresponding learning update, all agents know the (new) expressed opinions of all of their neighbors in the network. Hence, by fixing the dynamic transformations $\mathcal{I}$ and $\mathcal{R}$ and the observation $\mathcal{O}$ and the corresponding dynamic modalities $[\mathcal{I}][\mathcal{R}]$ and $[\mathcal{O}]$, we obtain a fragment of our general setting NKDL, the fragment $\mathsf{NKI^RO}$. Moreover, we will denote by KIO the (sub)fragment which leaves out the reflexive notion of social influence to contain only the simple influence.

To show how fragment $\mathsf{NKI^RO}$ captures what we wanted to model, let us now return the example of Figure 6.4. If we let $\mathcal{M}$ denote the model of Step 0 of Figure 6.4, then clearly the model of Step 1 is $\mathcal{M}^{\mathcal{I}}$ and the model of Step 1' is $(\mathcal{M}^{\mathcal{I}})^{\mathcal{O}}$. Applying the learning update amounts to cutting all accessibility links between any two worlds where a satisfies two different propositions from the set $\{ep, ec, en\}$. As a result, in Step 1' Bob comes to know Alice's private opinion simply because he knows her expressed opinion, that is, because he "observes" her change of expressed opinion.

Moreover, letting a and b be nominals for Alice and Bob respectively, and letting w be the actual world of Figure 6.4, it is not hard to see that $\mathcal{M}, w, b \models [\mathcal{I}][\mathsf{L}]K@_a ip$ (after a step of influence followed by a step of learning, b knows that a's inner opinion is pro), while $\mathcal{M}, w, a \models [\mathcal{I}][\mathsf{L}]\neg K@_b ip$ (after a step of influence followed by a step of learning a does not know that b's inner opinion is pro).

In summary, we have shown how our framework, and in particular fragment $\mathsf{NKI^RO}$, captures the type of examples discussed in the previous sections.[8]

6.4 Conclusion

6.4.1 "Opaque" Agents by Default

We started this chapter by considering the following limitation of our previous work: the notion of social influence we had specifically designed to capture influence among "opaque" agents actually sometimes lets agents "see through" each other. That is, as we have shown (in Section 6.2), agents who initially do not know each other's private opinion can come to "see through" each other as an effect of our "two-layer" notion of social influence. It seemed problematic, especially when remembering that we introduced "two-layered" agents to allow for distinguishing what agents can and cannot observe from each other, addressing our own objection to the "one-layer" model from [120]. And indeed, it seems strange that agents who know that their network neighbors privately agree with them would still feel *as strongly* to align on their expressed opinion.

So where do we stand now? Did the epistemization conducted by this chapter bring anything new on the table? Yes, it brought a more refined notion of social influence: when an agent a does not know her network neighbors' private opinions, she is influenced by their expressed opinions (exactly as before). However, should she come to know their private opinion (for any reason), then this information may

[8]We leave out of this chapter some of the work from our paper [64]. There, we provide a sound a complete tableau system for the "non-reflexive" fragment of the new logic (KIO), which also contains the first tableau system for the "Facebook Logic" from [145].

sometimes trump the information about their expressed opinion. More precisely, it trumps it only in cases where: 1) their expressed opinion is not sincere, and 2) their inner opinions disagree less with a's inner opinion than their expressed opinions. Hence, knowledge of the inner opinion of others is relevant only to the extent that it may lower the pressure on a to express an insincere opinion. Hence, we do not have to give up the claim made in Chapter 4, according to which the possibility of a mismatch between private mental states and observable behavior is essential to many emergent phenomena. The fact that agents sometimes know each other's mental state does not take away our point: what matters is that sometimes, they do *not* know it, and that a logic designed to capture social influence should allow for that possibility. The framework introduced in this chapter simply allows for *both* possibilities.

6.4.2 Summary

In this chapter, we have given a starting point for capturing the interplay between knowledge and diffusion phenomena:

- We have identified a possible objection to the work presented in Chapter 4: our "opaque" agents modeled in Chapter 4 can sometimes "see through" each other, i.e, come to know the others' private opinions simply by observing how their expressed opinion evolves – under the assumption of common knowledge of the rules of social influence.

- To reason about this type of learning induced by social influence, we have extended the hybrid diffusion logic (ND) from Chapter 5 with an epistemic dimension and some learning modalities, obtaining the dynamic hybrid epistemic logic NKDL for networks, knowledge, diffusion, and learning.

- We have modeled an extended example to show how ontic change (social influence) induces epistemic change (learning) which should affect the way agents are influenced by each other.

- We have proposed one way to take into account knowledge when modeling opinion diffusion induced by social influence, in an attempt to reconcile the ontic and epistemic correlated dynamics: we have proposed a new "reflexive" notion of social influence, according to which agents are less pressured to conform to the behavior of others around them when they know that this behavior does not reflect their private opinion sincerely.

- We have shown how the fragment $\mathsf{NKI^RO}$ of our logic is sufficient to model the cases of opinion diffusion considered so far.

- We have concluded that, while "smart" agents should indeed be able to see through the evolution of behavior and infer the private opinion of others,

> agents should still be assumed to be "opaque" by default: unless they know that the behavior of their neighbors does not reflect their private opinion, agents should still be influenced by their observable behavior.[9]

Note that our formal treatment of learning is actually more general than the examples discussed may have suggested. First, although we assumed that the network structure was common knowledge among the agents, our logic NKDL can also represent learning and reasoning about social influence even in cases where this assumption is dropped.[10] Also, our analysis shows quite generally how diffusion may increase the knowledge social neighbors have about each other. We have shown one way to use such higher knowledge as a precondition for diffusion dynamics, but clearly this is a general avenue of thought.

6.4.3 Further Research

The new "reflexive" (knowledge dependent) notion of social influence introduced in this chapter seems more adequate for modeling opinion change under social influence within intelligent agents than the simpler "two-layer" (knowledge-independent) notion of social influence introduced in Chapter 4. Overall, our new notion appears to be more appropriate with respect to the type of examples motivating this chapter, involving human agents who are typically "smart enough" to reason about each other's change of behavior.

However, we do not intend to claim that the latter is *the* correct notion of social influence. Our aim was to show that knowledge can be included in the definition of our dynamic transformations, and hence that our framework is flexible enough to accommodate many different rules of contagion, whether additional knowledge is relevant to the process or not.

Moreover, this chapter was restricted to take into account the first-order knowledge of agents. However, in situations involving more strategic interaction, this may not be sufficient, and higher-order knowledge may also need to be included in the preconditions of diffusion.

If an agent is "smart" enough to reason about higher-order knowledge, she should be able to reason about what her neighbors know that their neighbors know. If she also knows the rules of social influence, shouldn't she be able to predict how others will be influenced in the future? And if so, shouldn't she be influenced now by this knowledge of their future behavior?

[9]In addition to the material presented in this chapter, [64] contains some technical discussion that we do not include here. There, we give a complete terminating tableau system for a fragment (NKL) of our logic without dynamic transformations, which also happens to be the first tableau system for the Facebook logic from [145]. We also show how to reduce away the influence modality ($\mathcal{I}$) into that fragment, making the tableau for NKL a tableau system for NKIL and therefore for NKIO, capturing the "two-layer" influence from Chapter 4.

[10]To do so, one merely allows different network structures $\asymp_w$ at different worlds.

The next chapter answers this question by considering how agents with unbounded higher-reasoning powers can predict and anticipate diffusion. Hence, while this started by wondering how the dynamics of diffusion affect knowledge, the next chapter's starting point is this chapter's conclusion: how does knowledge affect diffusion dynamics?

Chapter 7

Knowledge as Diffusion Accelerator: DEL for Threshold Models

This chapter, based on [19], introduces a dynamic epistemic logic for diffusion phenomena. We model the diffusion process and show how knowing more about the other agents' behavior and about the network structure may interfere with such diffusion dynamics.

Three main steps are taken here. First, we focus on the dynamics of *threshold models*, as introduced in Section 2.1.2, where agents adopt a new behavior/product/opinion as soon as the proportion of their network-neighbors who have already adopted it meets the threshold, and do not "unadopt" it ever. In terms of the susceptible/infected model mentioned earlier, this amounts to agents who can catch a contagious infection, and stay infected forever.

Second, on the formal side, we leave behind the *hybrid* logic setting progressively enriched throughout Chapters 4, 5, and 6, to introduce an entirely new setting: a *propositional* dynamic epistemic logic for diffusion in threshold models. This allows us to reason both about networks and about knowledge without requiring a much more complex setting with two modal dimensions.

And third, while the previous chapter proposed a way to define knowledge-dependent diffusion rules, this was not pursued beyond first-order information. Here, we will move forward towards diffusion rules taking into account unbounded levels of higher-order reasoning. Hence, our agents become "smarter": we started with diffusion processes where agents reacted exclusively to their most direct environment, and end up with agents able to predict and anticipate diffusion, in order, for instance, to coordinate.

7.1 Introduction

The chapter has two goals. Our first goal is to design a logic for reasoning about threshold models and their dynamics. Our second goal is to investigate how the agents' knowledge affects such dynamics.

We construct a minimal dynamic propositional logic to describe the threshold dynamics and show that the logic is sound and complete.

Conceptually, the work presented here is in line with the earlier chapters in using logic to model diffusion phenomena within network structures. However, our new framework distinguishes itself by avoiding the use of static "neighborhood" modalities or hybrid logic tools. In this sense, the logical setting we introduce is "minimal": only propositional logic is used to specify both the network structure and the agents behavior, and a single dynamic modality is used to represent threshold-limited influence. Moreover, while our previous work focuses on the limit thresholds of 100% (all neighbors) and non-0% (at least one neighbor), we allow here for any (uniform) adoption threshold, as is standard within the literature on threshold models (see Section 2.1.2).

We then extend this framework with an epistemic dimension and investigate how information about more distant neighbors' behavior allows agents to anticipate changes in behavior of their closer neighbors. It is shown that this epistemic prediction dynamics is equivalent to the non-epistemic threshold model dynamics if agents know exactly their neighbors' behavior, but that more knowledge might *accelerate* diffusion, up to the point where maximal knowledge makes the whole diffusion jumps to its fixed point in only one step.

7.1.1 Outline

Section 7.2 introduces a minimal logic to capture the dynamics of diffusion in threshold models. We first recall the notion of threshold models introduced in Section 2.1.2, before defining a sound and complete dynamic propositional logic for modeling threshold influence within social networks. We then show how the logic captures the relationship between clusters and cascades.

In Section 7.3 we add knowledge to the picture. We introduce *epistemic* threshold models. These models come equipped with a knowledge-dependent update procedure, called "informed adoption", where agents must possess sufficient information about their surroundings before they adopt. This is a conceptual jump from the initial minimal modeling of influence from Section 7.2 to a more sophisticated (information dependent) diffusion policy: instead of agents adopting a behavior whenever enough of their neighbors have adopted it already, agents adopt whenever they *know* that enough of their neighbors have already adopted. We then relate these two adoption policies by showing under which epistemic conditions their diffusion dynamics are step-wise identical. The section is concluded by extending the logic to a sound and complete dynamic epistemic logic for the epistemic threshold models and the informed update procedure.

We further notice an interesting feature of the informed update procedure. Even though the "informed update" requires that agents have *enough* information to be influenced, the update does not require them to use *all* their available information when making their choices. Hence, if we consider threshold models as representing reflecting agents who are driven by a coordination goal, the new knowledge dependent update procedure makes our agents choose an action even when *they know they could do better*. To overcome this shortcoming, in Section 7.4, we introduce a third adoption policy, a "prediction update", where agents utilize *all* the available information to *predict the future behavior of other agents in the network*, and act upon their predictions. In other words, they *anticipate*, and it is common knowledge that they do. We show that the agents' reasoning about other predicting agents always reaches a fixed point and that making adoption dependent on this very fixed point captures the best response of agents trying to coordinate to the best of their knowledge. We give an example illustrating how knowledge about the network and about the behavior of other agents can be interpreted as an "accelerator" of diffusion dynamics, under this last prediction policy: the fixed point of the diffusion process under the prediction update is the same as under the informed update, but it can be reached faster if agents know more about the network around them. In Section 7.5 we discuss the relationship between how far agents can see and their prediction power.

Finally, Section 7.6 discusses the in-built assumptions of the introduced updates as well as some alternative diffusion policies and Section 7.7 gives some directions for further research.

7.1.2 Reaching for the Limits?

In the three previous chapters, we have assumed more and more complex agents, making them progressively "smarter". We have allowed them to have expressed opinions which do not reflect their private opinions, we have shown how knowledge can be gained as a consequence of diffusion and how this knowledge should be taken into account when modeling social influence.

Here, we go beyond this "one-step" improvements of agents, and beyond one-step dynamics, and we reach for limits: the limits of what knowledge agents with unbounded higher-order reasoning power have, and the limits of the diffusion processes. We will combine those limits to reach a result about maximally informed and maximally smart agents: when both the network structure and the agents' behavior are common knowledge, the diffusion dynamics reaches its speed limit: it jumps to its fixed point in only one step.

7.2 Threshold Models and their Dynamic Logic

This section designs a logic to capture the dynamics of threshold models. Section 7.2.1 first reminds the reader of the standard notion threshold models, previously introduced in Section 2.1.2.

7.2.1 Threshold Models for Social Influence

We start by recalling threshold models, introduced in 2.1, and formalizing our simplifying assumptions. As the previous ones, this chapter restricts itself to finite and undirected graphs without self-loops. We add here, for simplicity, that each agent has at least one neighbor in the network, as isolated agents are irrelevant to a discussion of social influence:

7.2.1. Definition. [Network] A network is a pair $(\mathcal{A}, N)$ where $\mathcal{A}$ is a non-empty finite set of *agents* and the function $N : \mathcal{A} \to \mathcal{P}(\mathcal{A})$ assigns a set $N(a)$ to each $a \in \mathcal{A}$, such that

$a \notin N(a)$	(Irreflexivity),
$b \in N(a)$ if and only if $a \in N(b)$	(Symmetry),
$N(a) \neq \emptyset$	(Seriality).

Recall that the simplest type of threshold model consists of such a network together with a unique behavior B (opinion, fashion, product, or "like-able item") distributed over $\mathcal{A}$ and a fixed uniform adoption threshold θ. A threshold model thus represents the current spread of behavior B throughout the network, while containing the adoption threshold which prescribes how this spread will evolve.

7.2.2. Definition. [Threshold Model]
A threshold model is a tuple $\mathcal{M} = (\mathcal{A}, N, B, \theta)$ where $(\mathcal{A}, N)$ is a network, $B \subseteq \mathcal{A}$ is a *behavior* and $\theta \in [0, 1]$ is a uniform *adoption threshold*.

As we have done so far, we keep assuming that the network structure stays unchanged. We also assume that the adoption threshold stay constant under updates. Thus, the spread of the behavior (i.e., the extension of B) at ensuing time steps may be calculated using the fixed threshold and network structure as follows:

7.2.3. Definition. [Threshold Model Update] The update of threshold model $\mathcal{M} = (\mathcal{A}, N, B, \theta)$ is the threshold model $\mathcal{M}' = (\mathcal{A}, N, B', \theta)$, where B' is given by

$$B' = B \cup \{a \in \mathcal{A} : \frac{|N(a) \cap B|}{|N(a)|} \geq \theta\}.$$

This definition, standard in the literature [71], captures the idea that the new set of agents who adopted the behavior B' (in the new updated model $\mathcal{M}'$) does include the set of agents B who had already adopted the behavior before and it includes those agents who have enough influential neighbors (given by the number θ) that have adopted already.

By repeatedly applying this update rule in an initial threshold model, we obtain a unique sequence of threshold models, which we call a diffusion sequence:

7.2.4. Definition. [Diffusion Sequence] Let $\mathcal{M} = (\mathcal{A}, N, B, \theta)$ be a threshold model. The diffusion sequence $\mathcal{S}_{\mathcal{M}}$ is the sequence of threshold models

$$\langle \mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, ..., \mathcal{M}_n, \mathcal{M}_{n+1}, ...\rangle,$$

such that, for any $n \in \mathbb{N}$, $\mathcal{M}_n = (\mathcal{A}, N, B_n, \theta)$ where B_n is given by:

$$B_0 = B \text{ and } B_{n+1} = B'_n.$$

Note that this diffusion process always reaches a fixed point, and that the number of agents in the model gives an upper bound on the number of updates that can be performed before reaching the fixed point:

7.2.5. Proposition. *Let $\mathcal{S}_{\mathcal{M}}$ be a diffusion sequence. For some $n \in \mathbb{N} < |\mathcal{A}|$, $\mathcal{M}_n = \mathcal{M}_{n+1}$.*

Proof: The fact that there is a $n \in \mathbb{N}$ such that $\mathcal{M}_n = \mathcal{M}_{n+1}$ follows immediately from the fact that $\mathcal{A}$ is finite and $B_n \subseteq B_{n+1}$ for all $n \in \mathbb{N}$. The fact that $n < |\mathcal{A}|$ is given by considering the slowest possible diffusion scenario, i.e. where $|B_0| = 1$ and only one agent adopts per round, i.e. for each $m < n \in \mathbb{N}$, $|B_m| = m + 1$. In this case $\left|B_{|\mathcal{A}|-1}\right| = |\mathcal{A}|$. □

Interpretation. Threshold models and their dynamics may be interpreted in two ways. One interpretation assumes that agents are mere automata and that their behavior is forced upon them by their environment. This interpretation suits the models that are used in e.g. epidemiology: viral infection "just happens" to agents. Alternatively, agents may be interpreted as rational beings aiming towards coordination with their neighbors. In fact, the above update rule also corresponds to the best response dynamics of an associated coordination game [126], under the assumption that there is a 'seed' set of players who always, possibly irrationally, play B [71].

Numerous variations of threshold models exist in the literature, including infinite networks [126], networks with non-inflating behavior adoption [126], agent-specific thresholds [109], weighted links [109] and multiple behaviors [8]. For simplicity, and to fit most examples in the literature, we will stick to the above simpler notion of finite threshold models. The next section proposes a logical framework to reason about them.

7.2.2 The Logic of Threshold-Limited Influence

This section introduces a minimal logic to model the standard notion of threshold-limited influence introduced in the section above. To describe the *situation* of a social network at a given moment, the static language needs to allow us to capture two things: who is related to whom and who is displaying the contagious behavior B. In this chapter, both features will be encoded using propositional variables. To describe the *change* of situation of a social network, the language includes a dynamic modality. This modality represents how agents adopt the behavior of their neighbors, whenever the given adoption threshold is reached, i.e., whenever enough neighbors have adopted.

7.2.6. DEFINITION. [Languages $\mathcal{L}_{[]}$ and $\mathcal{L}$] Let $\mathcal{A}$ be a finite set and let atoms be given by $\Phi = \{N_{ab} : a, b \in \mathcal{A}\} \cup \{\beta_a : a \in \mathcal{A}\}$. The language $\mathcal{L}_{[]}$ is then given by:

$$\varphi := N_{ab} \mid \beta_a \mid \neg\varphi \mid \varphi \wedge \varphi \mid [adopt]\varphi$$

The formulas of $\mathcal{L}$ are those of $\mathcal{L}_{[]}$ that do not involve the $[adopt]$-modality.

Disjunction and material implication are defined in the standard way. $\mathcal{L}_{[]}$ is an extension of propositional logic with a unary dynamic modality, denoted $[adopt]$. The language is interpreted over threshold models, using the behavior set and the social network to determine the extension of the atomic formulas. The $[adopt]$ modality is interpreted as is standard in Dynamic Epistemic Logic [24, 33, 70, 44]: we evaluate $[adopt]\varphi$ as true "today" if and only if φ is true "tomorrow". Here, "tomorrow" is given by the threshold update of Definition 7.2.3.

7.2.7. DEFINITION. [Truth Clauses for $\mathcal{L}_{[]}$] Given a model $\mathcal{M} = (\mathcal{A}, N, B, \theta)$, $N_{ab}, \beta_a \in \Phi$, and $\varphi, \psi \in \mathcal{L}_{[]}$:

$\mathcal{M} \models \beta_a$	iff	$a \in B$
$\mathcal{M} \models N_{ab}$	iff	$b \in N(a)$
$\mathcal{M} \models \neg\varphi$	iff	$\mathcal{M} \not\models \varphi$
$\mathcal{M} \models \varphi \wedge \psi$	iff	$\mathcal{M} \models \varphi$ and $\mathcal{M} \models \psi$
$\mathcal{M} \models [adopt]\varphi$	iff	$\mathcal{M}' \models \varphi$, where $\mathcal{M}'$ is the updated threshold model (Definition 7.2.3).

Let us also introduce some abbreviations:

7.2.8. DEFINITION. $[adopt]^n\varphi$ is defined recursively:

$$[adopt]^0\varphi := \varphi$$

$$[adopt]^{n+1}\varphi := [adopt][adopt]^n\varphi$$

7.2.9. DEFINITION. $\beta_{N(a)} \geq \theta$ expresses that the proportion of agent a's neighbors who have adopted is equal to or above the threshold θ:

$$\beta_{N(a)} \geq \theta := \bigvee_{\{\mathcal{G} \subseteq \mathcal{N} \subseteq \mathcal{A} : \frac{|\mathcal{G}|}{|\mathcal{N}|} \geq \theta\}} \Big(\bigwedge_{b \in \mathcal{N}} N_{ab} \wedge \bigwedge_{b \notin \mathcal{N}} \neg N_{ab} \wedge \bigwedge_{b \in \mathcal{G}} \beta_b\Big)$$

The following proposition captures within our language the fact (as noted in Prop. 7.2.5) that all diffusion sequences stabilize after some finite number of updates, illustrating how our language allows for capturing features of threshold model dynamics, such as stability and stabilization of the diffusion sequence:

7.2.10. PROPOSITION. *Let $\mathcal{M} = (\mathcal{A}, N, B, \theta)$ be a threshold model. There exists $n \in \mathbb{N} < |\mathcal{A}|$ such that, for any $\varphi \in \mathcal{L}_{[]}$:*

$$[adopt]^n \varphi \leftrightarrow [adopt]^{n+1} \varphi$$

Proof: As noted in the proof of Proposition 7.2.5, in the diffusion sequence $\mathcal{S}_{\mathcal{M}}$, for some $n \in \mathbb{N} < |\mathcal{A}|$, $\mathcal{M}_n = \mathcal{M}_{n+1}$.

Hence $\mathcal{M}_n$ and $\mathcal{M}_{n+1}$ are guaranteed to satisfy the same formulas, whereby $\mathcal{M} \models [adopt]^n \varphi \leftrightarrow [adopt]^{n+1} \varphi$. □

Axiomatization. We obtain an axiomatization of the logic for threshold models and their update dynamics by using the standard method of reduction rules from dynamic epistemic logic [24, 70, 33, 44].

7.2.11. DEFINITION. [The Logic of Threshold-Limited Influence, L_θ] The logic L_θ is comprised of any axiomatization of the propositional calculus and of the axioms and derivation rules of Table 7.1, for a given $\theta \in [0, 1]$.

The *static* logic consists of the axioms of propositional logic and the three network axioms of Table 7.1, and the rule of Modus Ponens. These capture the constraints imposed on the networks. In the *dynamic* part of the logic, we define rules that reduce formulas that contain the [*adopt*] modality to formulas without it. This is possible as the update procedure is deterministic: all the information required to determine the update threshold model is present in the current model. Hence the next state is "pre-encoded" in the present state.

As the [*adopt*] modality only affects the extension of B, the reduction axioms are trivial in all cases except those involving β_a. The corresponding reduction law, Red.Ax.β, relies on the mentioned pre-encoding. The axiom Red.Ax.β states that a has adopted B after the update just in case 1) she had already adopted it before the update or 2) the proportion of her neighbors who had already adopted it before the update was above threshold θ.

7.2.12. DEFINITION. [$\mathcal{C}_\theta$] Let $\theta \in [0, 1]$ be given. The class of threshold models $\mathcal{C}_\theta$ contains all and only models with the same threshold θ.

For any given threshold, the minimal logic L_θ is sound and complete with respect to the corresponding class of models $\mathcal{C}_\theta$:[1]

[1]The proof system and model class are further parametrized by the set of agents $\mathcal{A}$ used to define the corresponding language.

Network Axioms	
$\neg N_a a$	Irreflexivity
$N_{ab} \leftrightarrow N_b a$	Symmetry
$\bigvee_{b \in \mathcal{A}} N_{ab}$	Seriality
Reduction Axioms	
$[adopt]N_{ab} \leftrightarrow N_{ab}$	Red.Ax.N
$[adopt]\neg\varphi \leftrightarrow \neg[adopt]\varphi$	Red.Ax.$\neg$
$[adopt]\varphi \wedge \psi \leftrightarrow [adopt]\varphi \wedge [adopt]\psi$	Red.Ax.$\wedge$
$[adopt]\beta_a \leftrightarrow \beta_a \vee \beta_{N(a)} \geq \theta$	Red.Ax.β
Inference Rules	
From φ and $\varphi \rightarrow \psi$, infer ψ	Modus Ponens
From φ, infer $[adopt]\varphi$	$Nec_{[adopt]}$

Table 7.1: Hilbert-style proof system L_θ.

7.2.13. Theorem (Completeness). *Let $\theta \in [0,1]$. For any $\varphi \in \mathcal{L}$,*

$$\models_{\mathcal{C}_\theta} \varphi \textit{ iff } \vdash_{L_\theta} \varphi$$

Proof: *Soundness:* Let $\mathcal{M} = (\mathcal{A}, N, B, \theta)$ be an arbitrary threshold model with $a, b \in \mathcal{A}$. Then $\mathcal{M}$ satisfies Irreflexivity (Symmetry/seriality) directly by the semantics and the assumption of irreflexivity (symmetry/seriality) of the network. $\mathcal{M} \models [adopt]N_{ab} \leftrightarrow N_{ab}$ as the adoption operation never alters the network. Soundness of Red.Ax.$\neg$ and Red.Ax.$\wedge$ may be shown straightforwardly using induction on the length of formulas.

To see that $\mathcal{M}$ satisfies Red.Ax.β, let $\mathcal{M}'$ be the adoption update of $\mathcal{M}$. Then $\mathcal{M} \models [adopt]\beta_a$ iff $\mathcal{M}' \models \beta_a$ iff $a \in B' = B \cup \left\{b \in \mathcal{A} : \frac{N(b) \cap B}{N(b)} \geq \theta\right\}$ iff $\mathcal{M} \models \beta_a$ or $a \in \left\{b \in \mathcal{A} : \frac{N(b) \cap B}{N(b)} \geq \theta\right\}$. A syntactic decoding following Definition 7.2.9 of the large, right-hand disjunct of Red.Ax.β (called $\beta_{N(a) \geq \theta}$) shows that it is satisfied iff $a \in \left\{b \in \mathcal{A} : \frac{N(b) \cap B}{N(b)} \geq \theta\right\}$: The outer disjunction requires/ensures the existence of two sets of agents, $\mathcal{G}$ and $\mathcal{N}$, such that $\mathcal{G} \subseteq \mathcal{N}$ and $\frac{|\mathcal{G}|}{|\mathcal{N}|} \geq \theta$. The inner conjunction in Definition 7.2.9 is satisfied iff $\mathcal{N} = N(a)$ and $\mathcal{G} \subseteq B$. Hence φ is satisfied iff $\exists \mathcal{G} \subseteq N(a) \cap B : \frac{|\mathcal{G}|}{|N(a)|} \geq \theta$ iff $\frac{|N(a) \cap B|}{|N(a)|} \geq \theta$ iff $a \in \left\{b \in \mathcal{A} : \frac{N(b) \cap B}{N(b)} \geq \theta\right\}$. Hence $\mathcal{M} \models [adopt]\beta_a$ iff $\mathcal{M} \models \beta_a$ or $\mathcal{M} \models \beta_{N(a) \geq \theta}$.

Completeness: The proof goes via translation of the dynamic language into the static part of the language, in the usual way (see for instance [70, Ch. 7]). □

7.2.3 Clusters and Cascades

An agent adopting a new behavior may influence some of her neighbors to adopt it at the next moment, which in turn may cause further agents to adopt it, and so on. Recall that, as mentioned in Section 2.1, such a *cascade* is said to be *complete* when it results into a state where *all* agents have adopted the new behavior (see e.g. [71, Ch. 19]). Because the above given updates of threshold model always reach a fixed point, any cascade will eventually stop. However, a cascade may stop before all agents have adopted, i.e. without being complete. The following recalls a known result about how cascading effects are constrained by the network structure and shows how the suitable constraint may be captured by the minimal logic L_θ.

First of all, our language can express that a diffusion sequence will reach a complete cascade, given the upper bound on the number of updates before stabilization of the diffusion process noted in Proposition 7.2.5:

7.2.14. Definition. The sentence abbreviated by *cascade* expresses that all agents will have adopted eventually:

$$cascade := [adopt]^{|A|-1} \bigwedge_{a \in \mathcal{A}} \beta_a$$

As introduced in Section 2.1, strongly connected groups of agents are more resilient to external influence. Dense components of a network may prevent complete cascades and the denser a group, the better it resists change induced from the outside. The relevant notion of density is capture by the notion of a *cluster of density d* [71], as introduced in Section 2.1, defined by:

7.2.15. Definition. [Cluster of density d] Given a network $(\mathcal{A}, N)$ a cluster of density d is any group $C \subseteq \mathcal{A}$ such that for all $a \in C$,

$$\frac{|N(a) \cap C|}{|N(a)|} \geq d.$$

Example: clusters. Let model $\mathcal{M}$ given as illustrated below, with $B = \{d\}$. In this model, $C = \{a, b, c\}$ is a cluster of density $\frac{2}{3}$, in which no member belongs to B.

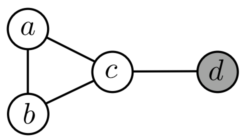


Figure 7.1: A social network with a cluster of density $\frac{2}{3}$.

The language $\mathcal{L}$ can express the existence of a cluster: if C is a cluster of density d then for each a in C, there is a big enough subset of C which are a's neighbors.

7.2.16. PROPOSITION. *The group C is a cluster of density d in $(\mathcal{A}, N)$ iff $\mathcal{M} = (\mathcal{A}, N, B, \theta)$ satisfies*

$$\bigwedge_{a \in C} \bigvee_{\left\{\mathcal{G} \subseteq \mathcal{N} \subseteq \mathcal{A} : \frac{|\mathcal{G} \cap C|}{|\mathcal{N}|} \geq d\right\}} \left(\bigwedge_{b \in \mathcal{N}} N_{ab} \wedge \bigwedge_{b \notin \mathcal{N}} \neg N_{ab} \right) \tag{7.1}$$

Proof: Left to right: Let $\mathcal{M} = (\mathcal{A}, N, B, \theta)$ and assume C is a cluster of density d in $(\mathcal{A}, N)$. Then by definition, for all $a \in C$, $\frac{|N(a) \cap C|}{|N(a)|} \geq d$. As $\mathcal{M}$ is based on $(\mathcal{A}, N)$, $\{b : \mathcal{M} \models N_{ab}\} = N(a)$ for all $a \in \mathcal{A}$. Let a be given and pick $\mathcal{N} = N(a)$ and $\mathcal{G} = N(a) \cap C$. Then $\frac{|\mathcal{G}|}{|\mathcal{N}|} \geq d$. Given the choice of $\mathcal{N}$, $\mathcal{M} \models \bigwedge_{b \in \mathcal{N}} N_{ab} \wedge \bigwedge_{b \notin \mathcal{N}} \neg N_{ab}$. So $\mathcal{M}$ satisfies (7.1).

Right to left: Assume that $\mathcal{M}$ satisfies (7.1) for some $C \subseteq \mathcal{A}$ and some $d \in [0, 1]$. Then for each $a \in C$, there is are sets $\mathcal{G}$ and $\mathcal{N}$ with $\mathcal{G} \subseteq \mathcal{N}$ and $\frac{|\mathcal{G} \cap C|}{|\mathcal{N}|} \geq d$, such that $\mathcal{N} = \{\mathcal{M} \models N_{ab}\} = N(a)$. Hence $\frac{|\mathcal{G} \cap C|}{|N(a)|} = \frac{|\mathcal{G} \cap C|}{|\mathcal{N}|} \geq d$. As $\mathcal{G} \cap C \subseteq \mathcal{N} = N(a)$, $\frac{|N(a) \cap C|}{|N(a)|} \geq d$. As a was arbitrary from C, C is indeed a cluster of density d in $(\mathcal{A}, N)$. □

Alternative proof: As $\mathcal{A}$ is assumed finite, (7.1) can be unfolded to a finite, propositional formula equivalent with the second-order statement

$$\forall a \in C \exists \mathcal{G}' \subseteq C : \left(\frac{|\mathcal{G}'|}{|N(a)|} \geq d \wedge \forall b \in \mathcal{G}' : b \in N(a) \right). \tag{7.2}$$

(7.2) states that for each agent a in C, there is a sufficiently large (relative to d and the size of a's neighbourhood) subgroup $\mathcal{G}'$ of C, such that all of $\mathcal{G}'$ is in the neighbourhood of a. Hence (7.2) states that for all $a \in C$, a has a large enough proportion of neighbors in C for C to be a cluster.

Given Proposition 7.2.16, it is easy to see that the sentence below characterizes the existence of a cluster of density d among agents *who have not adopted* (abbreviated $\exists C_{\geq d} \neg \beta$):

$$\exists C_{\geq d} \neg \beta := \bigvee_{C \subseteq \mathcal{A}} \bigwedge_{a \in C} \bigvee_{\left\{\mathcal{G} \subseteq \mathcal{N} \subseteq \mathcal{A} : \frac{|\mathcal{G} \cap C|}{|\mathcal{N}|} \geq d\right\}} \left(\bigwedge_{b \in \mathcal{N}} N_{ab} \wedge \bigwedge_{b \notin \mathcal{N}} \neg N_{ab} \wedge \bigwedge_{b \in \mathcal{G}} \neg \beta_b \right)$$

Note that we can express in the same way that there is a cluster of density greater than d, by replacing $\geq$ by the strict $>$ in the formula (abbreviated $\exists C_{>d} \neg \beta$).

Example: clusters, cont. The model illustrated in Figure 7.1 contains a cluster $C = \{a, b, c\}$ of density $\frac{2}{3}$, such that no agent in C has adopted. Hence, the model should satisfy $\exists C_{\frac{2}{3}} \neg\beta$:

$$\bigvee_{C \subseteq \mathcal{A}} \bigwedge_{a \in C} \bigvee_{\left\{\mathcal{G} \subseteq \mathcal{N} \subseteq \mathcal{A} : \frac{|\mathcal{G} \cap C|}{|\mathcal{N}|} \geq \frac{2}{3}\right\}} \left(\bigwedge_{b \in \mathcal{N}} N_{ab} \wedge \bigwedge_{b \notin \mathcal{N}} \neg N_{ab} \wedge \bigwedge_{b \in \mathcal{G}} \neg\beta_b \right). \tag{7.3}$$

To verify this, assume C is a group that satisfies the outmost disjunction. Then for each $a \in C$ there is must a $\mathcal{G}$ and $\mathcal{N}$ such that $\frac{|\mathcal{G} \cap C|}{|\mathcal{N}|} \geq \frac{2}{3}$ for which $\mathcal{M}$ satisfies

$$\bigwedge_{b \in \mathcal{N}} N_{ab} \wedge \bigwedge_{b \notin \mathcal{N}} \neg N_{ab} \wedge \bigwedge_{b \in \mathcal{G}} \neg\beta_b. \tag{7.4}$$

To see that $\mathcal{M}$ satisfies (7.4), regard first agent c, for whom the appropriate $\mathcal{N}$ is $\{a, b, d\}$. As $|\mathcal{N}| = 3$, we must identify a group $\mathcal{G} \subseteq C$ with $|\mathcal{G}| \geq 2$ such that for all $b \in \mathcal{G}$, $\mathcal{M} \models N_{cb}$. Such a $\mathcal{G}$ exists, being $\{a, b\}$. Finally, indeed $\mathcal{M} \models \neg\beta_a \wedge \neg\beta_b$, and hence the conjunct for c is satisfied. Similar reasoning shows that the conjuncts for a and b also hold. This gives us (7.3).

The Cluster Theorem. Let us recall here the theorem from [126],[71, Ch.19.3], introduced in Section 2.1.2 characterizing the possibility of a complete adoption cascade in a network:

> Given a threshold model $\mathcal{M}$ with threshold $\theta \neq 0$ and a set $B \subset \mathcal{A}$ of agents who have adopted, all agents will eventually adopt *if, and only if* there does not exist a cluster of density greater than $1 - \theta$ in $\mathcal{A} \backslash B$.

As both the complete cascade and the existence of the relevant clusters are expressible in $\mathcal{L}_{[]}$, the cluster theorem can also be encoded in our setting, in the following way:

Let $\mathcal{M} = (\mathcal{A}, N, B, \theta)$ with $\theta \neq 0$. Then

$$\mathcal{M} \models \textit{cascade} \leftrightarrow \neg\exists C_{>1-\theta} \neg\beta.$$

7.2.4 Logics for Generalizations of Threshold Models

So far, we have considered the "simplest" possible network structures: the networks are finite, symmetric, irreflexive and serial. The constraints of symmetry and irreflexivity could easily be relaxed in the initial definition of threshold models (Def. 7.2.2) to generalize the logics to different types of social relationships (for instance a hierarchical network).

For simplicity, we work with uniform thresholds. Obtaining logics for settings without this uniformity constraint is unproblematic: 1) define θ not as a constant but as a function assigning a particular threshold to each agent; i.e., set $\theta : \mathcal{A} \to [0, 1]$ in the definition of threshold models (Def. 7.2.2); 2) replace θ by $\theta(a)$ in the definition of the update (Def. 7.2.3) and in the reduction axiom Red.Ax.β (in Table 7.1). This will generate a logic for each such function θ, that is, for each distribution of thresholds among agents.

The logical setting may also be generalized to capture the spread of *several* behaviors and their interaction. This amounts to: 1) modify the definition of threshold models (Def. 7.2.2) to let $\mathcal{B}$ be a finite *set* of behaviors ($\mathcal{B} = \{B_1, B_2, ...B_n\}$) and define $\theta : \mathcal{A} \times \mathcal{B} \to [0, 1]$; 2) Relativize the definition of the update to each behavior B_i; 3) extend our set of atomic propositions: $\Phi = \{N_{ab} : a, b \in \mathcal{A}\} \cup \{\beta_{ia} : a \in \mathcal{A}, i \in 1, ...n\}$; 4) relativize the semantic clause in the obvious way: $\mathcal{M} \vDash \beta_{ia}$ iff $a \in B_i$, and replace the reduction axiom $Red.Ax.\beta$ by $Red.Ax.\beta_i$ accordingly. The "signature" of the resulting logic would be given by $[\theta, \mathcal{A}, \mathcal{B}]$. Such a logic allows reasoning about the diffusion of a fixed number of behaviors, given a specific distribution of thresholds for each behavior to each agent, for any particular network structure.

Furthermore, we consider the *proportion* of neighbors who have adopted the only relevant factor for decision making. This makes every neighbor as influential as any other. To generalize, weighted links representing different "degrees of influence" could be used instead. The condition for being influenced into adopting would become: the *weighted sum* of my neighbors which have adopted is at least θ. Alternatively, we could fix an ordering of neighbors of each agent a with $b \geq_a c$ stating that agent b influences agent a at least as much as agent c does. Based on such an ordering, on possible update policy would be that a adopts when a given proportion of $\geq_a$-maximal agents have adopted.

Additional alternative policies will be discussed in Section 7.6. These will also involve epistemic considerations, the topic to which we turn next.

7.3 Epistemic Threshold Models and Their Dynamic Logic

By the definition of the above given standard update on threshold models, agents *react to their environment*: they are always influenced by the actual behavior of their direct neighbors. In many situations, this "nomothetic" update style seems to pose unrealistic requirements. The update requires that agents act in accordance with the *facts* of others' behavior, even in the face of uncertainty. Hence, the standard update may require of agents that they act in accordance with information that they do not possess. For an example, see Figure 7.2.

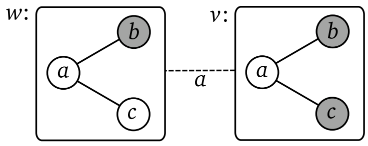


Figure 7.2: A situation of uncertainty. Agent a cannot tell whether world w or world v is the actual one, as indicated by the dashed line (when representing indistinguishability relations we omit reflexive and transitive links). Hence, a does not know whether c has adopted or not. Assume that the threshold $\theta > \frac{1}{2}$ and that v is the actual world. Then, according to the standard update, a should adopt – but a does not know that!

To accommodate this shortcoming, we extend the standard threshold models with an epistemic dimension and define a refined adoption policy where agents' behavior change depends on their knowledge of others' behavior. We moreover define a logic suitable to reason about *epistemic threshold models* and their dynamics.

To add an epistemic dimension to threshold models, we add for each agent a subjective epistemic indistinguishability relation in the standard way, as illustrated in Figure 7.2.

7.3.1 Epistemic Threshold Models

The most general version of threshold models with an epistemic dimension that we will work with in this chapter is the following:

7.3.1. Definition. [Epistemic Threshold Model (ETM)] An epistemic threshold model (ETM) is a tuple $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$, where:

$\mathcal{W}$	is a finite, non-empty set of possible worlds (or states),
$\mathcal{A}$	is a finite non-empty set of agents,
$\sim_a \subseteq \mathcal{W} \times \mathcal{W}$	is an equivalence relation, for each agent $a \in \mathcal{A}$,
$N : \mathcal{W} \to (\mathcal{A} \to \mathcal{P}(\mathcal{A}))$	assigns a neighborhood $N(w)(a)$ to each $a \in \mathcal{A}$ in each $w \in \mathcal{W}$, such that: $a \notin N(w)(a)$ (Irreflexivity) $b \in N(w)(a) \Leftrightarrow a \in N(w)(b)$ (Symmetry) $N(w)(a) \neq \emptyset$ (Seriality)
$B : \mathcal{W} \to \mathcal{P}(\mathcal{A})$	assigns to each $w \in \mathcal{W}$ a set $B(w)$ of agents who have adopted.
$\theta \in [0, 1]$	is a uniform adoption threshold.

To reason about the impact of knowledge on diffusion in network situations, we want to impose limiting assumptions regarding the agents' uncertainty. It is for example natural to assume that agents know who their direct neighbors are, though cases exist where it is natural that agents know more about the network. Agents may know who the neighbors of neighbors are, or maybe the whole network is even common knowledge. Likewise, the uncertainty about agents' behavior might be subject to various constraints: agents may know the behavior of their neighbors, of their neighbors' neighbors, of everybody, etc.

One way to impose restrictions on uncertainty is by giving agents an egocentric "sphere of sight", corresponding to how far they can "see" in the network, assuming that if they can see further, they can see closer. We will say that an agent has *sight n* when she can "see" *at least* n agents away, i.e., when she knows *at least both* the network structure and the behavior of all agents within n distance. To provide a formal definition, we first fix what is meant by within "n distance":

7.3.2. Definition. [n-reachable] Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ and let $n \in \mathbb{N}$. Define $N^n : \mathcal{W} \to \mathcal{A} \to \mathcal{P}(\mathcal{A})$ as follows, for any $w \in \mathcal{W}$ and any $a, b, c \in \mathcal{A}$:

- $N^0(w)(a) = \{a\}$
- $N^{n+1}(w)(a) = N^n(w)(a) \cup \{b \in \mathcal{A} : \exists c \in N^n(w)(a) \text{ and } b \in N(w)(c)\}$

If $b \in N^n(w)(a)$, then b belongs to the set of agents that a has within her sight at world w. Morever, if $b \in N^n(w)(a)$ we say that b is *n-reachable* from a in w.

7.3.3. Definition. [Sight n Model[2]]
A sight n ETM $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ is such that, for $n \in \mathbb{N}$ and for any $a, b \in \mathcal{A}$, and any $w, v \in \mathcal{W}$:

- If $w \sim_a v$ and $b \in N^{n-1}(w)(a)$, then $N(w)(b) = N(v)(b)$ (agents know the network at least up to distance n)
- If $w \sim_a v$ and $b \in N^n(w)(a)$, then $b \in B(w)$ iff $b \in B(v)$ (agents know the behavior of others at least up to distance n).

7.3.2 Knowledge-Dependent Diffusion

To remedy the problem of agents acting on information they may not possess, we introduce a revised adoption policy. It captures the intuitive idea that an agent should only be influenced by what he knows about other agents around him. This amounts to a knowledge-dependent adoption policy: agents adopt whenever *they know that* enough of their neighbors have adopted already. We call this update policy *informed update*:

[2]We lump two notions of sight under one heading. A more general definition would be of sight (n, m), where n specifies the sight of network structure, while m specifies sight of behavior.

7.3.4. Definition. [Informed Update] Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be an ETM with sight n. The informed adoption update of $\mathcal{M}$ produces ETM $\mathcal{M}^i = (\mathcal{W}, \mathcal{A}, N, B^i, \theta, \{\sim_a^i\}_{a\in\mathcal{A}})$ such that, for any $a, b \in \mathcal{A}$ and any $w, v \in \mathcal{W}$:

- $B^i(w) = B(w) \cup \{a \in \mathcal{A} : \forall v \sim_a w \frac{|N(v)(a)\cap B(v)|}{|N(v)(a)|} \geq \theta\}$ and
- $w \sim_a^i v$ iff i) $w \sim_a v$ and ii) if $b \in N^n(w)(a)$, then $b \in B^i(w)$ iff $b \in B^i(v)$.

The first condition tells us that the new set of adopters at world w includes the previous set of adopters $B(w)$ (hence agents do not give up their previously adopted behavior) and it includes also all agents who, as far as they know, are certain of the fact that enough influential neighbors (given by θ) have adopted already. The second condition ensures that the informed update of an ETM with sight n is again an ETM with sight n, i.e., agents' sight is not diminished by updates.

Updating de dicto and updating de re. The informed update policy is defined using *de dicto* knowledge of others' behavior: if an agent knows that enough others will adopt, so should she, ignoring that she might not know exactly *who* will adopt.

A *de re* update is definable by setting

$$B^i(w) = B(w) \cup \{a \in \mathcal{A} : \frac{b \in \mathcal{A} : \forall v \sim_a w, |N(v)(a) \cap B(v)|}{|N(v)(a)|} \geq \theta\}.$$

While both rules are interesting, in the remainder of this chapter we opt for the *de dicto* version as it expresses in a stronger sense that agents can fully utilize all their information while staying in the spirit of threshold models.

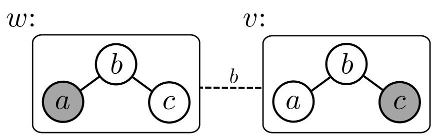


Figure 7.3: Adoption *de re* vs. adoption *de dicto*. We illustrate an ETM with threshold $\theta = \frac{1}{2}$ and two possible worlds. Should b adopt or not? He knows *de dicto* that enough neighbors have adopted, but he does not know so *de re*; he knows that at least half of his neighbors have adopted, but he doesn't know *which* half.

Learning the distribution. We have discussed in the previous chapter how ontic change may induce learning. We can make a similar observation here: when performing informed updates, agents may learn about the initial distribution of behavior in the network even outside their range of sight, as it may be possible to exclude possibilities based on the development of the dynamics. The learning occurs due to the restriction of the indistinguishability relation, as build into the definition of informed update. Figure 7.4 provides an example.

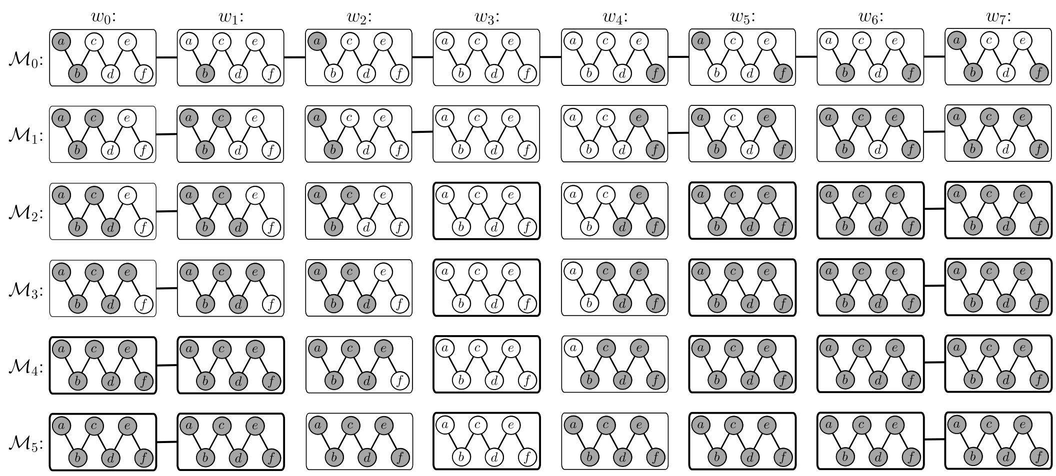


Figure 7.4: The learning process for agent d (bottom center) under blind adoption, in an ETM with threshold $\theta \leq \frac{1}{2}$ and sight 1. With sight 1, the ETM contains the 8 depicted possible worlds/states. The last states to reach fixed points at time 5 are states w_2 and w_4 from the left. Epistemic relations are drawn only for d to simplify representation. Note the development of the indistinguishability relation from $\mathcal{M}_0$ to $\mathcal{M}_5$: as the updated $\sim'_d$ is a restriction of $\sim_d$ to states where both c and e's behaviors are identical, d learns about the initial distribution. Learning may or may not be complete: compare the development of states w_1 and w_2.

Implicit information and redundant knowledge. Under some epistemic conditions, the epistemic and non-epistemic diffusion policies are equivalent. If each agent always knows *at least* who her neighbors are and how they are behaving, then the two policies give rise to the same diffusion dynamics, in the following sense: the diffusion dynamics resulting from the informed update on an ETM reduces to the diffusion dynamics under the initial (non-epistemic) updated applied to each possible world of the ETM. This is the content of Proposition 7.3.6 below.

Proposition 7.3.6 relates two important insights. The first is that standard threshold models make the *implicit epistemic assumption* that agents know their neighborhood and its behavior. The second is that *knowledge about more distant agents is redundant* as it will not affect behavior.

To prove the result, we first define how to generate a (non-epistemic) threshold model from a possible state of an epistemic threshold model:

7.3.5. Definition. [State-Generated Threshold Model (SGM)]
Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be an ETM and let $w \in \mathcal{W}$ and $a \in \mathcal{A}$. The state-generated threshold model $\mathcal{M}(w) = (\mathcal{A}, N_{\mathcal{M}(w)}, B_{\mathcal{M}(w)}, \theta)$ is given by:

$$N_{\mathcal{M}(w)}(a) = N(w)(a), \quad \text{and}$$
$$a \in B_{\mathcal{M}(w)} \Leftrightarrow a \in B(w).$$

7.3.6. Proposition. *Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be an ETM and $w \in \mathcal{W}$. Let $\mathcal{M}^i$ and $\mathcal{M}(w)$ be respectively the informed update and state-generated models of $\mathcal{M}$. Let $\mathcal{M}^i(w)$ be the state-generated model of $\mathcal{M}^i$ and let $\mathcal{M}(w)'$ non-epistemic update of $\mathcal{M}(w)$. Then*

$$\textit{if } \mathcal{M} \textit{ has sight } n \geq 1, \textit{ then } \mathcal{M}^i(w) = \mathcal{M}(w)'.$$

Proof: As neither the non-epistemic update nor the informed update changes the set of agents, network or threshold, it need only be shown that $B^i(w) = B(w)'$ where $B^i(w)$ is the behavior set of $\mathcal{M}^i(w)$ and $B(w)'$ is the behavior set of $\mathcal{M}(w)'$.

Assume $a \in B(w)$. Then $a \in B(w)^i$ from $\mathcal{M}^i$ by monotonicity of informed update. Hence also $a \in B_{\mathcal{M}^i(w)}$ from $\mathcal{M}^i(w)$ by Definition 7.3.5 of SGMs. From $a \in B(w)$ it also follows that $a \in B_{\mathcal{M}(w)}$ by defintion of SGMs. By monotonicity of the non-epistemic update, $a \in B'_{\mathcal{M}(w)}$ from $\mathcal{M}(w)'$.

Assume that $a \notin B(w)$. Then $a \notin B_{\mathcal{M}}(w)$ by definition of SGMs.

By definition, $a \in B(w)^i$ iff $\forall v \sim_a w : \frac{|N(v)(a)|\cup B(v)}{|N(v)(a)|} \geq \theta$. As $\mathcal{M}$ has sight $n \geq 1$, $\forall v \sim_a w \; N(v)(a) = N(w)(a)$ and $b \in N(w)(a)$ implies $b \in B(w) \Leftrightarrow b \in B(v)$. Hence $\frac{|N(w)(a)|\cup B(w)}{|N(w)(a)|} \geq \theta$. As $N(w)(a) = N_{\mathcal{M}(w)}(a)$ and $B(w) = B_{\mathcal{M}(w)}$, it follows that $\frac{|N_{\mathcal{M}(w)}(a)|\cup B_{\mathcal{M}(w)}}{|N_{\mathcal{M}(w)}(a)|} \geq \theta$ iff $a \in B_{\mathcal{M}(w)}$. □

Proposition 7.3.6 provides a precise, but partial, interpretation of the dynamics of non-epistemic threshold models as an information-dependent behavior diffusion. As witnessed by its proof, only the immediate neighborhood of agents matters for the adoption behavior in a threshold model. This changes when agents are imbued with predictive abilities; see Section 7.4.

The interpretation is only partial in that we *do not* obtain a full characterization of the standard threshold dynamics (see Definition 7.2.3) by requiring sight $n \geq 1$. Sight $n < 1$ *does not* imply that there will always be a *difference making* neighbor about which some agent has uncertainty. If a has uncertainty about some neighbor b's behavior but is certain that a large enough proportion of neighbors have adopted, then the model will have sight strictly less than 1 while still developing according to the standard threshold dynamics.

Situations in which neighbors lack knowledge of some direct neighbors' behavior are interesting in that they may cause the diffusion process to *slow down* compared to the standard update policy:

7.3.7. Proposition. *There exists an ETM $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ with sight $n < 1$ such that*

$$B_{\mathcal{M}^i(w)} \subset B_{\mathcal{M}(w)^u},$$

where $\mathcal{M}^i$ and $\mathcal{M}(w)$ are respectively the informed update and state-generated models of $\mathcal{M}$, and $\mathcal{M}^i(w)$ is the state-generated model of $\mathcal{M}^i$ and $\mathcal{M}(w)'$ is the non-epistemic update (Def. 7.2.3) of $\mathcal{M}(w)$.

Proof: By construction of a specific model: Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ with $W = \{w, v\}$, $w \sim_a v$, $N(w)(a) = N(v)(a)$ but $\frac{|N(w)(a)\cap B(w)|}{|N(w)(a)|} \geq \theta > \frac{|N(v)(a)\cap B(v)|}{|N(v)(a)|}$. Then $a \notin B_{\mathcal{M}^i(w)}$, but $a \in B_{\mathcal{M}(w)'}$. □

Figure 7.5 illustrates this "slower" diffusion.

7.3.3 Knowledge and Cascades

In Section 7.2.3, we have shown how our language can capture complete cascades and the existence of clusters able to block diffusion, as captured by the *Cluster Theorem*: a cascade will be complete if and only if the network does not contain a cluster of non-yet-adopters of density greater than $1 - \theta$.

Given proposition 7.3.6 above, the cluster theorem still holds for any epistemic threshold model with sight at least 1. Moreover, the existence of a relevant cluster will still block a cascade under the informed update policy, independently of how much agents know. However in general, considering any epistemic threshold model with any sight, the cluster theorem cannot be maintained as it was stated. What we observe is that the left to right direction of the cluster theorem still holds for epistemic threshold models with sight less than 1: indeed, if a complete cascade occurs, then the network does not contain a cluster of density greater than $1 - \theta$.

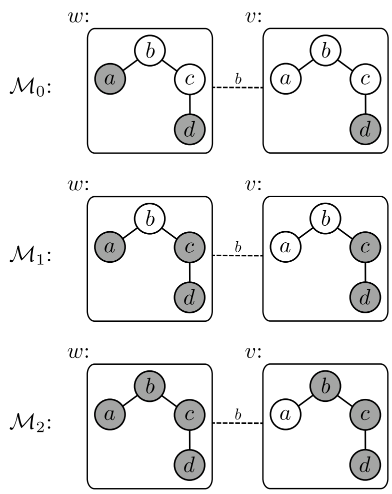


Figure 7.5: A diffusion process "slowed down" by the uncertainty of agent b, with threshold $\theta = \frac{1}{2}$. Consider the situation in world w: agent a has adopted, but agent b does not know it. Therefore, agent b will not adopt immediately. The diffusion according to the informed update policy in state w will only stabilize after applying the informed update rule *twice*. Note that under the non-epistemic threshold update, or if agent b knew whether a has adopted, the situation depicted in w would stabilize after only one step (i.e. the non-epistemic threshold update of $\mathcal{M}_0(w)$ gives us directly $\mathcal{M}_2(w)$).

However, the converse of does not hold in these models with sight less than 1. We briefly explain this point in more detail. Given proposition 7.3.7 above, we know that the diffusion process, via the informed update rule, in an ETM with sight < 1 might be "slower" than the process based on the non-epistemic threshold update policy. Indeed, the lack of knowledge may for instance block a cascade, despite the absence of a cluster-obstacle. Figure 7.6 illustrates this difference.

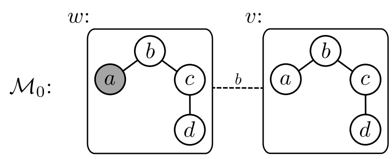


Figure 7.6: A diffusion process "blocked" by the uncertainty of agent b, with $\theta = \frac{1}{2}$. Consider the situation in world w: agent a has adopted, but agent b does not know it. Therefore, agent b will not adopt (under the informed adoption rule). Note that under the non-epistemic threshold update, or if agent b knew that a has adopted, the situation depicted in state w would evolve into a complete cascade.

7.3.4 The Epistemic Logic of Threshold-Limited Influence

To reflect the epistemic dimension in a formal syntax, the language $\mathcal{L}$ is extended by adding the standard K_a modalities reading "agent a knows that", for each agent $a \in \mathcal{A}$.

7.3.8. DEFINITION. [Languages $\mathcal{L}_{K[]}$ and $\mathcal{L}_K$] Let the set of atomic propositions be given by $\{N_{ab} : a, b \in \mathcal{A}\} \cup \{\beta_a : a \in \mathcal{A}\}$ for a finite set $\mathcal{A}$. Where $a, b \in \mathcal{A}$, the formulas of $\mathcal{L}_{K[]}$ are given by

$$\varphi := N_{ab} \mid \beta_a \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_a\varphi \mid [adopt]\varphi$$

The formulas of $\mathcal{L}_K$ are those of $\mathcal{L}_{K[]}$ that do not involve the $[adopt]$ modality.

As standard, we use the given language to define the other Boolean operators for disjunction and implication and introduce $< adopt >$ as the dual of $[adopt]$.

7.3.9. DEFINITION. [Semantics for $\mathcal{L}_{K[]}$ with Informed Update] Formulas $\varphi, \psi \in \mathcal{L}_{K[]}$ are interpreted over an ETM $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ with sight n, $w, v \in \mathcal{W}$:

$\mathcal{M}, w \models \beta_a$	iff	$a \in B(w)$
$\mathcal{M}, w \models N_{ab}$	iff	$b \in N(w)(a)$
$\mathcal{M}, w \models \neg\varphi$	iff	$\mathcal{M}, w \not\models \varphi$
$\mathcal{M}, w \models \varphi \wedge \psi$	iff	$\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
$\mathcal{M}, w \models K_a\varphi$	iff	for all $v \in \mathcal{W}$ such that $v \sim_a w$, $\mathcal{M}, v \models \varphi$
$\mathcal{M}, w \models [adopt]\varphi$	iff	$\mathcal{M}', w \models \varphi$, where $\mathcal{M}'$ is the informed update of $\mathcal{M}$ as specified in Def. 7.3.4.

Axiomatization. In the specification of the epistemic reduction axioms, the following two syntactic shorthands are used:

Abbreviation. For any $k \in \mathbb{N} \geq 1$, we introduce the abbreviation N^k_{ab} by induction,

$$\begin{aligned} N^1_{ab} &:= N_{ab} \\ N^{k+1}_{ab} &:= N^k_{ab} \vee \bigvee_{c\in\mathcal{A}} \left(N^k_{ac} \wedge N_{cb}\right) \end{aligned}$$

The formula N^k_{ab} expresses that b is k-reachable from a.

Network Axioms	
$\neg N_{aa}$	Irreflexivity
$N_{ab} \leftrightarrow N_{ba}$	Symmetry
$\bigvee_{b \in \mathcal{A}} N_{ab}$	Seriality
Knowledge Axioms	
$K_a\varphi \to \varphi$	$(*)$ Ax.T
$K_a\varphi \to K_aK_a\varphi$	$(*)$ Ax.4
$\neg K_a\varphi \to K_a\neg K_a\varphi$	$(*)$ Ax.5
Reduction Axioms	
$[adopt]N_{ab} \leftrightarrow N_{ab}$	Red.Ax.N
$[adopt]\neg\varphi \leftrightarrow \neg[adopt]\varphi$	Red.Ax.$\neg$
$[adopt]\varphi \wedge \psi \leftrightarrow [adopt]\varphi \wedge [adopt]\psi$	Red.Ax.$\wedge$
$[adopt]\beta_a \leftrightarrow \beta_a \vee K_a(\beta_{N(a)} \geq \theta)$	$(*)$ Ep.Red.Ax.β
$[adopt]K_a\varphi \leftrightarrow \bigvee_{\mathcal{B} \subseteq \mathcal{A}} (\mathcal{B} = N_a^n\beta^+ \wedge K_a (\mathcal{B} = N_a\beta^+ \to [adopt]\varphi))$	$(*)$ Ep.Red.Ax.K sight n
Inference Rules	
From φ and $\varphi \to \psi$, infer ψ	Modus Ponens
From φ, infer $K_a\varphi$ for any $a \in \mathcal{A}$	$(*)$ Nec.K_a
From φ, infer $[adopt]\varphi$	Nec.$[adopt]$

Table 7.2: Axioms and rules for the Epistemic Logic of Threshold-Limited Influence for sight n. Subscripts a, b are arbitrary over $\mathcal{A}$. Entries marked $(*)$ are new or modified relative to Table 7.1.

Abbreviation. For $\mathcal{B} \subseteq \mathcal{A}$, we introduce the abbreviation $\mathcal{B} = N_a^k\beta^+$ as follows:

$$\left(\mathcal{B} = N_a^k\beta^+\right) := \bigwedge_{b \in \mathcal{B}} \left(N_{ab}^k \wedge [adopt]\beta_b\right) \wedge \bigwedge_{b \in \mathcal{A} \setminus \mathcal{B}} \left(N_{ab}^k \to [adopt]\neg\beta_b\right).$$

The expression $\mathcal{B} = N_a^k\beta^+$ refers to the set of agents which are 1) k-reachable from a and 2) will have adopted after the next update.

Using these shorthands, the axioms for Epistemic Threshold Models and the dynamics of Informed Update are as given in Table 7.2.

The reduction law *Ep.Red.Ax.β* states that a has adopted β after the update just in case she had already adopted it before the update, or *she knew that* she had a large enough proportion of neighbors who had already adopted it before the update. *Ep.Red.Ax.K.sight.n* captures that an agent knows that φ will be the case after the update if, and only if, *she knows that*, if those very agents who actually are going to adopt do adopt, then φ will hold after the update.

7.3.10. Definition. [Epistemic Logic of Threshold-Limited Influence] The logic $L_{\theta n}$ is comprised of the axioms and rules of propositional logic and the axioms and rules of Table 7.2.

7.3.11. Definition. [$\mathcal{C}_{\theta n}$] Let $\theta \in [0, 1]$ be given. The class of ETM $\mathcal{C}_{\theta n}$ contains all and only ETM with threshold θ and sight n.

The logic $L_{\theta n}$ is sound and complete with respect to the corresponding class of models $\mathcal{C}_{\theta n}$:

7.3.12. Theorem. *Let $\theta \in [0, 1]$ and $n \in \mathbb{N}$. For any $\varphi \in \mathcal{L}_{K[]}$,*

$$\models_{\mathcal{C}_{\theta n}} \varphi \textit{ iff } \vdash_{L_{\theta n}} \varphi.$$

Proof: *Soundness:* Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be an epistemic threshold model with sight n. Let $a, b \in \mathcal{A}$ and $w, v \in \mathcal{W}$. Then $(\mathcal{M}, w)$ satisfies the S5 axioms as all $\sim_a$ are equivalence relations and satisfy the axioms reoccuring from Table 7.1 for the same reasons non-epistemic threshold models satisfy them.

To see that $(\mathcal{M}, w)$ satisfies Ep.Red.Ax.β, let $\mathcal{M}^i$ be the informed update of $\mathcal{M}$. Then $\mathcal{M}, w \models [adopt]\beta_a$ iff $\mathcal{M}^i, w \models \beta_a$ iff $a \in B^i(w) = B(w) \cup \left\{b \in \mathcal{A} : \forall v \sim_b w \frac{|N(v)(b)\cap B(v)|}{|N(v)(b)|} \geq \theta\right\}$ iff $\mathcal{M}, w \models \beta_a$ or $a \in \left\{b \in \mathcal{A} : \forall v \sim_b w \frac{|N(v)(b)\cap B(v)|}{|N(v)(b)|} \geq \theta\right\}$. Using the same syntactic decoding as in the proof of Theorem 7.2.13, we obtain that $a \in \left\{b \in \mathcal{A} : \forall v \sim_b w \frac{|N(v)(b)\cap B(v)|}{|N(v)(b)|} \geq \theta\right\}$ iff $\mathcal{M}, w \models K_a\left(\beta_{N(a)} \geq \theta\right)$. Hence $\mathcal{M}, w \models [adopt]\beta_a$ iff $\mathcal{M}, w \models \beta_a$ or $\mathcal{M}, w \models K_a\left(\beta_{N(a)} \geq \theta\right)$.

For Ep.Red.Ax.K.sight.n, let again $\mathcal{M}^i$ be the informed update of $\mathcal{M}$. Then

$$\mathcal{M}, w \models \bigvee_{\mathcal{B}\subseteq\mathcal{A}} ((\mathcal{B} = N_a^n\beta^+) \quad\wedge\quad K_a((\mathcal{B} = N_a\beta^+) \to [adopt]\varphi))$$
$$\text{iff}$$
$$\exists\mathcal{B} \subseteq \mathcal{A} : \mathcal{M}, w \models (\mathcal{B} = N_a^n\beta^+) \quad\wedge\quad K_a((\mathcal{B} = N_a\beta^+) \to [adopt]\varphi)$$
$$\text{iff}$$
$$\exists\mathcal{B} \subseteq \mathcal{A} : \mathcal{M}, w \models \bigwedge_{b\in\mathcal{B}} (N_{ab}^n \wedge [adopt]\beta_b) \wedge \bigwedge_{b\in\mathcal{A}\setminus\mathcal{B}} (N_{ab}^n \to [adopt]\neg\beta_b) \text{ and}$$
$$\mathcal{M}, w \models K_a\left(\left(\bigwedge_{b\in\mathcal{B}} (N_{ab}^n \wedge [adopt]\beta_b) \wedge \bigwedge_{b\in\mathcal{A}\setminus\mathcal{B}} (N_{ab}^n \to [adopt]\neg\beta_b)\right) \to [adopt]\varphi\right)$$
$$\text{iff}$$
$$\exists\mathcal{B} \subseteq \mathcal{A} : \mathcal{B} = N^n(w)(a) \cap B^i \text{ and}$$
$$\text{for all } v \sim_a w, \text{ if } \mathcal{B} = N^n(v)(a) \cap B^i, \text{ then } \mathcal{M}^i, v \models \varphi$$
$$(*)$$
$$\text{iff}$$

$$\exists \mathcal{B} \subseteq \mathcal{A} : \mathcal{B} = N^n(w)(a) \cap B' \text{ and}$$
$$\text{if } \mathcal{B} = N^n(w)(a) \cap B^i, \text{ then } \mathcal{M}^i, w \models K_a\varphi$$
$$\text{(from } (*) \text{ as } \mathcal{M} \text{ is sight } n, \text{ so } N^n(v)(a) \cap B^i = N^n(w)(a) \cap B^i \text{ for all } v \sim_a w)$$
$$\text{iff}$$
$$\mathcal{M}^i, w \models K_a\varphi$$
$$\text{(as such a } \mathcal{B} \text{ always exists)}$$
$$\text{iff}$$
$$\mathcal{M}, w \models [adopt]K_a\varphi.$$

Completeness (sketch): It can be shown by induction that for all $\varphi \in \mathcal{L}_{K[]}$, there exists a $\varphi' \in \mathcal{L}_K$ such that $\vdash_{L_{n\theta}} \varphi \leftrightarrow \varphi'$. Completeness then follows from the standard proof of completeness for S5 over Kripke models with equivalence relations and the straightforward insight that the network axioms characterize the imposed network conditions. □

7.4 Prediction Update

In defining our informed update rule based on epistemic threshold models, we ensure that agents do not act on information they do not possess. Such agents are however still limited, in that they do not take *all* their available information into account. This section investigates effects of agents that are allowed to reason about more than only the *present* behavior of the network. In particular, we focus on providing agents with *predictive power.*

Consider the ETM illustrated in Figure 7.7, with a given dynamics that runs according to a blind or informed adoption policy.

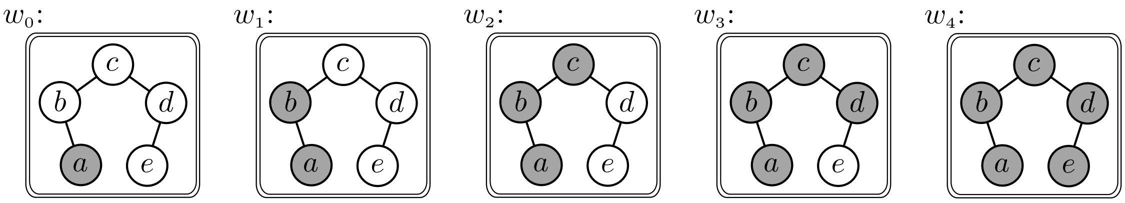


Figure 7.7: An ETM with no uncertainty about the actual state w, developing according to informed update. B is marked by gray, and a threshold $\theta = \frac{1}{2}$ is assumed. At time 0 (w_0), only a has adopted. According to informed adoption, b adopts at time 1. At time 2, c also adopts the behavior, etc.

If one assumes that agents (nodes) are not merely blindly influenced by their neighbors, but rather are rational agents seeking to coordinate, the dynamics in Figure 7.7 seems to misfire. In particular, as the network and behavior distribution are known to c (and if the new behavior is considered the most valuable), the choice of c *not* to adopt during the first update is irrational. As c knows that a has adopted, he knows that b will adopt during the next update round. Hence c also knows that he will be better off in round 1 if he, too, has chosen to adopt. To represent this "predictive rationality" we define a new, predictive, update mechanism.

Prediction update as the least fixed point. In defining "prediction update", we make use of the notion of a *least fixed point*. Even when agents' attempt to use all their available information, each will at some point reach a conclusion about her next action. When the last agent does so, the prediction reaches a fixed point.

This fixed point may be approximated using a chain of lower level predictions. The intuitive idea of the approximation may be illustrated using Figure 7.7:

> Assume agent a considers himself smart by predicting that he knows his only neighbor b is going to adopt B in round 2, if b follows the informed update policy. Then a may act preemptively, by also adopting B in round 2, rather than in round 3 as the informed update prescribes.[3] In this case, a may be thought of as a *level 1 predictor*: he assumes no-one else makes predictions, that the others are of level 0. However, a may come to think that b is as smart as he is, i.e., that also b is a level 1 predictor. Assuming this, a now foresees that b will not adopt in round 2, but already in round 1; based on this prediction about b's predictions, a may now also adopt in round 1. In this case, a is a *level 2 predictor*, etc.

If this reasoning is pushed to its fixed point, it will "catch up with itself": in the fixed point, every agent will be a level ω predictor, predicting under the assumption that all others are the same. This is the trick we use to ensure that agents draw the most powerful conclusion available.

Common knowledge of predictive rationality and complete information use. Prediction update incorporates two epistemic assumptions. One is that it is common knowledge that all agents act in accordance with the prediction update policy. This assumption means that agents may not only predict the systems behavior as if everybody else was acting in accordance with informed update. Rather, agents foresee the behavior of other predictors.

Moreover, it is common knowledge that predictors predict as far into the future as possible, given their information. This means that predictors attempt to use all their available information about the network structure, the current behavior spread and information available to others when determining their next action.

[3] If a acted according to the informed update policy, he must first see b adopt before he is influenced by b's choice

Prediction update preliminaries. Before we define the prediction update, a few preliminaries are required.

7.4.1. Definition. [Functions Γ_g] Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be a finite[4] ETM and let the set of all functions from $\mathcal{W}$ to $\mathcal{P}(\mathcal{A})$ be denoted by $\mathcal{P}(\mathcal{A})^{\mathcal{W}} = \{f | f : \mathcal{W} \to \mathcal{P}(\mathcal{A})\}$.

For each $g \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$ let the function $\Gamma_g : \mathcal{P}(\mathcal{A})^{\mathcal{W}} \longrightarrow \mathcal{P}(\mathcal{A})^{\mathcal{W}}$ be given by $\forall w \in \mathcal{W}, \forall f \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$

$$\Gamma_g(f)(w) = g(w) \cup \left\{ a \in \mathcal{A} : \forall v \sim_a w, \frac{|N(v)(a) \cap f(v)|}{|N(v)(a)|} \geq \theta \right\}.$$

7.4.2. Lemma. *Let $\mathcal{M}$, $\mathcal{P}(\mathcal{A})^{\mathcal{W}}$ and Γ_g be as in Definition 7.4.1. Let $\preceq$ be a partial order on $\mathcal{P}(\mathcal{A})^{\mathcal{W}}$ such that for any $f, g \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$, all $w \in \mathcal{W}$, $f \preceq g \Leftrightarrow f(w) \subseteq g(w)$. Then*

1) *$(\mathcal{P}(\mathcal{A})^{\mathcal{W}}, \preceq)$ is a finite, hence complete, lattice.*

2) *For each $g \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$, the map Γ_g is order-preserving (monotonic).*

Proof:

1) For any finite set $\mathcal{A}$, $(\mathcal{P}(A), \subseteq)$ is a finite and hence complete lattice with the order given by the set-theoretic inclusion. If $(L, \sqsubseteq)$ is a finite lattice and $\mathcal{W}$ a finite set, then $(L^{\mathcal{W}}, \leq)$ is also a finite lattice when $L^{\mathcal{W}} = \{f | f : \mathcal{W} \longrightarrow L\}$ and $f \leq g$ iff $\forall w \in \mathcal{W}, f(w) \sqsubseteq g(w)$. Hence, given that $\mathcal{W}$ is a finite set, also $(\mathcal{P}(\mathcal{A})^{\mathcal{W}}, \preceq)$ is a finite lattice with the order given by definition of $\preceq$. Every lattice over a finite set is also complete.

2) Let $g, f, f' \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$, and let $f \preceq f'$. Hence $\forall w \in \mathcal{W}$, $f(w) \subseteq f'(w)$. Pick an arbitrary $u \in \mathcal{W}$. Then

$$\begin{aligned}
\Gamma_g(f)(u) =& g(u) \cup \left\{ a \in \mathcal{A} : \forall v \sim_a u, \frac{|N(v)(a) \cap f(v)|}{|N(v)(a)|} \geq \theta \right\} \\
\Gamma_g(f')(u) =& g(u) \cup \left\{ a \in \mathcal{A} : \forall v \sim_a u, \frac{|N(v)(a) \cap f'(v)|}{|N(v)(a)|} \geq \theta \right\}.
\end{aligned}$$

Let the second terms of the unions be denoted A and A', respectively.

For all $v \in \mathcal{W}$, as $f(v) \subseteq f'(v)$, $\frac{|N(v)(a)\cap f(v)|}{|N(v)(a)|} \geq \theta$ implies $\frac{|N(v)(a)\cap f'(v)|}{|N(v)(a)|} \geq \theta$. Hence $A \subseteq A'$, so $\Gamma_g(f)(u) \subseteq \Gamma_g(f')(u)$. As u was arbitrary, $\Gamma_g(f) \preccurlyeq \Gamma_g(f')$. Hence Γ_g is order-preserving. As g was arbitrary, this holds for all $\Gamma_g, g \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$.
□

[4] In a finite ETM we assume that the set of worlds $\mathcal{W}$ is finite and the set of agents $\mathcal{A}$ is finite.

7.4.3. Definition. [Least Fixed Point] Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be a finite ETM and let $(\mathcal{P}(\mathcal{A})^{\mathcal{W}}, \preceq)$ be as in Lemma 7.4.2. Let Γ_g be as in Definition 7.4.1.

The *least fixed point* of Γ_g, $\texttt{lfp}(\Gamma_g)$, is the unique $x \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$ such that

$$\Gamma_g(x) = x, \text{ and}$$
$$\forall y \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}, \text{ if } \Gamma_g(y) = y, \text{ then } x \preceq y.$$

7.4.4. Theorem (lfp Existence). *Let $\mathcal{M}, (\mathcal{P}(\mathcal{A})^{\mathcal{W}}, \preceq)$ and Γ_g be as in Definition 7.4.3. Then $\texttt{lfp}(\Gamma_g)$ exists.*

Proof: The least fixed point $\texttt{lfp}(\Gamma_g)$ exists by the Knaster-Tarski Fixed Point Theorem (see e.g. [67, p. 50]), as $(\mathcal{P}(\mathcal{A})^{\mathcal{W}}, \preceq)$ is a complete lattice (Lemma 7.4.2) and Γ_g is order-preserving (Lemma 7.4.2). □

Defining prediction update. Given the previous paragraph, we may now define prediction update as follows:

7.4.5. Definition. [Prediction Update] Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be a finite ETM of sight n and let $(\mathcal{P}(\mathcal{A})^{\mathcal{W}}, \preceq)$ be as in Lemma 7.4.2. Let $\Gamma_B : \mathcal{P}(\mathcal{A})^{\mathcal{W}} \longrightarrow \mathcal{P}(\mathcal{A})^{\mathcal{W}}$ be given as per Definition 7.4.1, i.e., the function such that $\forall w \in \mathcal{W}, \forall f \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$

$$\Gamma_B(f)(w) = B(w) \cup \left\{ a \in \mathcal{A} : \forall v \sim_a w, \frac{|N(v)(a) \cap f(v)|}{|N(v)(a)|} \geq \theta \right\}.$$

The prediction update of $\mathcal{M}$ results in the ETM $\mathcal{M}' = (\mathcal{W}, \mathcal{A}, N, \widetilde{B}, \theta, \{\sim'_a\}_{a\in\mathcal{A}})$ where $\forall w, v \in \mathcal{W}$,

$$\widetilde{B}(w) = \texttt{lfp}(\Gamma_B)(w), \text{ and}$$
$$w \sim'_a v \text{ iff } w \sim_a v \text{ and if } b \in N^{\leq n}(w)(a), \text{ then } b \in \widetilde{B}(w) \text{ iff } b \in \widetilde{B}(v).$$

Finding the prediction update fixed point. The definition of prediction update does not provide us with a method for finding the least fixed point. The following theorem guarentees that we can find it using a bottom-up method:

7.4.6. Theorem. *Let $\mathcal{M}, (\mathcal{P}(\mathcal{A})^{\mathcal{W}}, \preceq)$ be as in Lemma 7.4.2 with bottom element $\perp$. Let Γ_B and $\widetilde{B}$ be defined as in Definition 7.4.5. Then*

$$\texttt{lfp}(\Gamma_B) = \sup\{{\Gamma_B}^n(\perp) : n \in \mathbb{N}\}$$

Proof: This proof follows from the proof of the Knaster-Tarski Fixed Point Theorem applied to finite structures. Given that we work with a finite structure $(\mathcal{P}(\mathcal{A})^{\mathcal{W}}, \preceq)$ and that Γ_B is order-preserving, a least fixed point is reached in a constructive way in finitely many steps. The construction is similar to Proposition 3.1. of [111]. $\square$

The above stated prediction update rule in definition 7.4.5 can now be used to give a new semantics to the [*adopt*] modality in the logic language $\mathcal{L}_{K[]}$.

7.4.7. Definition. [Semantics for $\mathcal{L}_{K[]}$ with Prediction Update] Given a finite ETM $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ with sight n, $w \in \mathcal{W}$, and $\varphi \in \mathcal{L}_{K[]}$ truth clauses are as in Definition 7.3.9, except for $\varphi := [adopt]\psi$, $\psi \in \mathcal{L}_{K[]}$ given by

$$\mathcal{M}, w \models [adopt]\varphi \text{ iff } \mathcal{M}', w \models \varphi, \text{ where } \mathcal{M}' \text{ is the prediction update of } \mathcal{M}.$$

Axiomatization. We provide sound axioms that govern the least fixed point behavior of the prediction update policy, but we do not provide a complete axiom system. Finding a complete logic is the aim of planned future research. For now we introduce a fixed point axiom and a least fixed point rule of inference. Note that in this section, the [*adopt*] modality is a fixed point operator and hence may no longer be reduced away. Contrary to the informed update process, using prediction update results in a system that is strictly more expressive than its non-dynamic counterpart.

To state the proof system, we first generalize the syntactic shorthand introduced in Definition 7.2.9.

Abbreviation. Given a tuple of formula's $(\varphi_b)_{b\in\mathcal{A}}$, one for each agent $a \in \mathcal{A}$, we introduce the following abbreviation:

$$K_a(\varphi_{N(a)} \geq \theta) := K_a\left(\bigvee_{\left\{\mathcal{G}\subseteq\mathcal{N}\subseteq\mathcal{A}:\,\frac{|\mathcal{G}|}{|\mathcal{N}|}\geq\theta\right\}}\left(\bigwedge_{b\in\mathcal{N}} N_{ab} \wedge \bigwedge_{b\notin\mathcal{N}} \neg N_{ab} \wedge \bigwedge_{b\in\mathcal{G}} \varphi_b\right)\right).$$

Here $K_a(\varphi_{N(a)} \geq \theta)$ denotes that a knows that larger than a θ fraction of her neighbors has the property φ (where for instance φ_b can stand for $N_{ab} \wedge \beta_b$). In particular, $K_a([adopt]\beta_{N(a)} \geq \theta)$ expresses that a knows that at least a θ fraction of her neighbors will have adopted β after the application of the prediction update rule.

7.4.8. Definition. [Prediction Logic] The logic $L^{predict}_{\theta n}$ is comprised of the axioms and rules of propositional logic and the axioms and rules of Table 7.2 with the only change that the axiom *Ep.Red.Ax.*β is replaced by the Fixed Point Axiom in Table 7.3 and we extend the set of rules of the logic with the least fixed point inference rule in Table 7.3.

Fixed Point Laws	
$[adopt]\beta_a \leftrightarrow \beta_a \vee K_a([adopt]\beta_{N(a)} \geq \theta)$	Fixed Point Axiom
$\dfrac{\vdash \{\varphi_a \leftrightarrow \beta_a \vee K_a(\varphi_{N(a)} \geq \theta)\}_{a\in\mathcal{A}}}{\vdash \{\varphi_a \to [adopt]\beta_a\}_{a\in\mathcal{A}}}$	Least Fixed Point Inference Rule

Table 7.3: Fixed point laws of the prediction logic $L_{\theta n}^{predict}$.

The Fixed Point axiom of Table 7.3 is almost identical to Ep.Red.Ax.β of Table 7.2, except for the inclusion of the [*adopt*] modality on the right-hand side. This states that a will adopt after the prediction update iff she has already adopted, or if she knows that enough of her neighbors will have adopted *after having applied the same predictive reasoning she uses.*

The Least Fixed Point Inference rule reflects the fact that prediction update is defined as a least fixed point operator.

We do not provide a complete logic for the prediction update policy. It is our conjecture that the axioms and rules in definition 7.4.8 will be necessary to obtain completeness. The listed axioms and rules are sound with respect to epistemic threshold models using the prediction update rule as our semantics for the [*adopt*] modality. For the axioms and rules not in Table 7.3, the proof of Theorem 7.3.12 carries over. The axiom and rule governing the fixed point behavior is shown to be sound in the following proposition.

7.4.9. Proposition. *The axiom and derivation rule of Table 7.3 are sound with respect to epistemic threshold models with sight n, using prediction update as semantics for the* [adopt] *modality.*

Proof: Let $\mathcal{M}$ be a arbitrary finite ETM with sight n, domain $\mathcal{W} \ni w$ and $a, b \in \mathcal{A}$.

Fixed Point Axiom. $\mathcal{M}, w \models [adopt]\beta_a$ iff $\mathcal{M}', w \models \beta_a$ iff $a \in \widetilde{B} = B \cup \left\{b \in \mathcal{A} : \forall v \sim_b w, \frac{|N(v)(b)\cap\widetilde{B}|}{|N(v)(b)|} \geq \theta\right\}$ iff $\mathcal{M}, w \models \beta_a$ or $\forall v \sim_a w, \frac{|N(v)(a)\cap\widetilde{B}|}{|N(v)(a)|} \geq \theta$. The right disjunct obtains iff

$$\forall v \sim_a w, \exists \mathcal{G}, \mathcal{N} \subseteq \mathcal{A} : \quad \begin{array}{l} \mathcal{G} \subseteq \mathcal{N} \text{ and } \frac{|\mathcal{G}|}{|\mathcal{N}|} \geq \theta \text{ and} \\ \mathcal{G} \subseteq \widetilde{B} \text{ and } \mathcal{N} = N(v)(a) \end{array}$$

iff

$$\forall v \sim_a w, \exists \mathcal{G}, \mathcal{N} \subseteq \mathcal{A} : \begin{array}{l} \mathcal{G} \subseteq \mathcal{N} \text{ and } \frac{|\mathcal{G}|}{|\mathcal{N}|} \geq \theta \text{ and} \\ \forall b \in \mathcal{G},\ \mathcal{M}', v \models \beta_b \text{ and } \forall b \in \mathcal{N}, \\ \mathcal{M}', v \models N_{ab} \end{array}$$

iff

$$\forall v \sim_a w,\ \mathcal{M}', v \models \bigvee_{\left\{\mathcal{G} \subseteq \mathcal{N} \subseteq \mathcal{A} : \frac{|\mathcal{G}|}{|\mathcal{N}|} \geq \theta\right\}} \left(\bigwedge_{b \in \mathcal{N}} N_{ab} \wedge \bigwedge_{b \notin \mathcal{N}} \neg N_{ab} \wedge \bigwedge_{b \in \mathcal{G}} \beta_b \right)$$

iff

$$\forall v \sim_a w,\ \mathcal{M}, v \models \bigvee_{\left\{\mathcal{G} \subseteq \mathcal{N} \subseteq \mathcal{A} : \frac{|\mathcal{G}|}{|\mathcal{N}|} \geq \theta\right\}} \left(\bigwedge_{b \in \mathcal{N}} N_{ab} \wedge \bigwedge_{b \notin \mathcal{N}} \neg N_{ab} \wedge \bigwedge_{b \in \mathcal{G}} [adopt]\beta_b \right)$$

iff

$$\mathcal{M}, w \models K_a \left(\bigvee_{\left\{\mathcal{G} \subseteq \mathcal{N} \subseteq \mathcal{A} : \frac{|\mathcal{G}|}{|\mathcal{N}|} \geq \theta\right\}} \left(\bigwedge_{b \in \mathcal{N}} N_{ab} \wedge \bigwedge_{b \notin \mathcal{N}} \neg N_{ab} \wedge \bigwedge_{b \in \mathcal{G}} [adopt]\beta_b \right) \right)$$

iff

$$\mathcal{M}, w \models K_a([adopt]\beta_{N(a)} \geq \theta)$$

Hence we conclude $\mathcal{M}, w \models [adopt]\beta_a$ iff $\mathcal{M}, w \models \beta_a \vee K_a([adopt]\beta_{N(a)} \geq \theta)$.

Least Fixed Point Inference Rule. Let an arbitrary finite ETM $\mathcal{M}$ with sight n and domain $\mathcal{W}$ be given. Where $\{\varphi_a\}_{a \in \mathcal{A}}$ is a set of sentences from $\mathcal{L}_{K[]}$, let $\overline{\varphi} \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$ with $\overline{\varphi}(w) = \{a \in \mathcal{A} : \mathcal{M}, w \models \varphi_a\}$. Moreover, let $\Gamma_{\overline{\varphi}} : \mathcal{P}(\mathcal{A})^{\mathcal{W}} \longrightarrow \mathcal{P}(\mathcal{A})^{\mathcal{W}}$, given by

$$\Gamma_{\overline{\varphi}}(f) = h \text{ such that}$$
$$\forall w \in \mathcal{W}, h(w) = \overline{\varphi}(w) \cup \left\{ a \in \mathcal{A} : \forall v \sim_a w, \frac{|N(v)(a) \cap f(v)|}{|N(v)(a)|} \geq \theta \right\}.$$

As shown in Lemma 7.4.2, each such $\Gamma_{\overline{\varphi}}$ is order-preserving.

Let $\overline{\beta} \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$ be determined by $\{\beta_a\}_{a \in \mathcal{A}}$ and $\overline{[]\beta} \in \mathcal{P}(\mathcal{A})^{\mathcal{W}}$ by $\{[adopt]\beta_a\}_{a \in \mathcal{A}}$. Let $\Gamma_{\overline{\beta}}$ be given by the above construction.

Given the prediction semantics of $[adopt]$ and the fact that $\widetilde{B} = \mathtt{lfp}(\Gamma_B) = \sup\{\Gamma_B{}^n(\bot) : n \in \mathbb{N}\}$ (Theorem 7.4.6), we may conclude that

$$\overline{[]\beta} = \Gamma_{\overline{\beta}}(\overline{[]\beta}) \tag{7.5}$$

is the least fixed point of $\Gamma_{\overline{\beta}}$.

Assume for some $\{\varphi_a\}_{a\in\mathcal{A}}$ that $\vdash \left\{\varphi_a \leftrightarrow \beta_a \vee K_a(\varphi_{N(a)} \geq \theta)\right\}_{a\in\mathcal{A}}$. This implies

$$\vdash \bigwedge_{a\in\mathcal{A}} (\varphi_a \leftrightarrow \beta_a \vee K_a(\varphi_{N(a)} \geq \theta)). \tag{7.6}$$

From $\{\varphi_a\}_{a\in\mathcal{A}}$ and $\{\beta_a \vee K_a(\varphi_{N(a)} \geq \theta)\}_{a\in\mathcal{A}}$ we may define functions $\overline{\varphi}$ and $\overline{\beta K}$, as specified above. Now notice that $\overline{\beta K} = \Gamma_{\overline{\beta}}(\overline{\varphi})$. Hence, for (7.6) to be satisfied, we have that

$$\overline{\varphi} = \Gamma_{\overline{\beta}}(\overline{\varphi}).$$

Given that (7.5) is the least fixed point of $\Gamma_{\overline{\beta}}$, we have that $\overline{\varphi} = \Gamma_{\overline{\beta}}(\overline{\varphi}) \Rightarrow \overline{[]\beta} \preceq \overline{\varphi}$, so

$$\begin{array}{lll} \forall w \forall a: & a \in \overline{[]\beta}(w) \Rightarrow a \in \overline{\varphi}(w) & \text{so} \\ \forall w \forall a: & w \models [adopt]\beta_a \Rightarrow w \models \varphi_a & \text{so} \\ \forall w \forall a: & w \models [adopt]\beta_a \rightarrow \varphi_a & \end{array}$$

□

7.5 Sight and Prediction Power

Relationship between predictive power and non-epistemic update. Similarly, as we compared the informed update policy with the non-epistemic threshold model update, it is also natural to investigate the relationship between the 'prediction update', 'informed update' and the 'non-epistemic threshold model update' (Definition 7.2.3). Indeed, given that the prediction update policy foresees the non-epistemic deterministic development of the actual state under uncertainty, such a comparison would be rather natural. Besides comparing the cascading behavior and speed of convergence, (as illustrated in figure 7.5), other results that we expect in this investigation relate to posing conditions and finding a lower and upper bound of how far agents can predict into the future. We leave the technical details of this investigation for future work.

Bounded rationality. Stating that prediction update is the fixed point of the informed update, as we have done in this section, corresponds to assuming that agents have unbounded rationality (maximal anticipation power given the information available). A *bounded* rationality version of the prediction update dynamics could be defined, in which agents can only anticipate a fixed finite number of steps ahead. A natural way of doing this would be by defining an update that updates to some finite level n of the prediction chain. The dynamics of bounded rationality would only differ from the unbounded dynamics for a low enough n. We leave the full exploration of technical details of the prediction update involving such boundedly rational agents for future work.

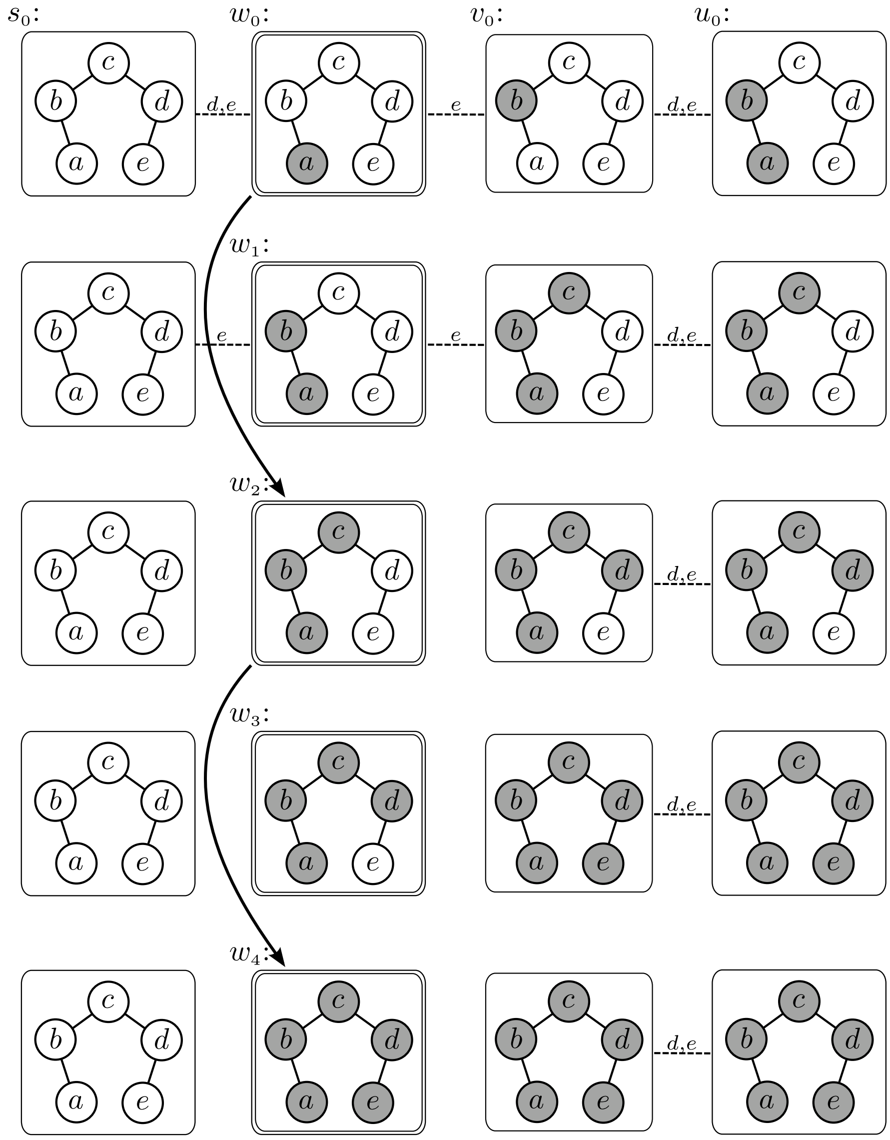


Figure 7.8: We use the prediction update to regulate the dynamics of this sight 2, finite ETM with actual state w, $\theta = \frac{1}{2}$ (reflexive and transitive uncertainty relations are omitted in the illustration). Agents a, b, c are endowed with additional information: they are fully informed about the actual state. The development of the states is given according to blind/informed adoption; states w_0–w_4 are from Figure 7.7. The thick arrow indicates the evolution of the actual world under the specified prediction dynamics. With informed update, w reaches a fixed point after 4 updates; with prediciton update, it reaches the same fixed point after only 2 steps. Due to uncertainty, the prediction update does not jump to the fixed point of the non-epistemic update in 1 step: as d does not know whether a has adopted at time 0, she does not know that c will adopt under the prediction update. Hence, she will refrain herself from adopting until w_3. Similar considerations goes for e.

7.6 Alternative Adoption Policies

In the previous sections, we have presented three diffusion policies: one depending solely on whether the agents' neighbors have adopted (the "threshold model update" from Def. 7.2.3); one depending on *knowledge* of this fact (the "informed update" of Def. 7.3.4), and one depending on the *anticipation* of this fact (the "prediction update" of Def. 7.4.5). This section questions some in-built assumptions of these policies and discusses possible alternatives.

Enlarging the sphere of influence. The adoption policies hitherto presented rely on the idea that an agent will adopt if (she knows that) enough of her *direct neighbors* (will) have adopted.

For certain applications, decisions are made that are based not only on actions of direct neighbors, but on the population at large. A case in point is the decision of whether to *support a revolution*: the relevant factor is then whether a big enough fraction of the total population supports the revolution, not whether enough of one's direct neighbors do so.

Generally, such policies may be obtained by enlarging the "sphere of influence" of agents beyond their direct neighbors. A natural choice in the epistemic setting is to fit the "sphere of influence" to agents' "sphere of sight" (in models of sight n). The influence principles would then become: the agent adopts if (he knows that) enough of his *n-distant* neighbors (will) have adopted.

In the revolution case, agents might be influenced into adopting only if (they know that) enough agents *within the whole network* (will) have adopted. A suitable "globalized" version of the prediction update from Def. 7.4.5 may be defined as follows:

7.6.1. Definition. [Global Prediction Update]

Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be a sight n finite model, and let $(F, \leq)$ be as in Def. 7.4.5. The *global prediction update* of $\mathcal{M}$ results in the model $\mathcal{M}' = (\mathcal{W}, \mathcal{A}, N, \widetilde{B}, \theta, \{\sim'_a\}_{a\in\mathcal{A}})$ where:

- $\widetilde{B}$ is such that:
  - $\forall w \in \mathcal{W}, \widetilde{B}(w) = B(w) \cup \{a \in \mathcal{A} : \forall v \sim_a w, \frac{|\mathcal{A}\cap\widetilde{B}(v)|}{|\mathcal{A}|} \geq \theta\}$
  - $\forall f \in F$, if $\forall w \in \mathcal{W}, f(w) = B(w) \cup \{a \in \mathcal{A} : \forall v \sim_a w, \frac{|\mathcal{A}\cap f(v)|}{|\mathcal{A}|} \geq \theta\}$, then $\widetilde{B} \leq f$.

  and
- $w \sim'_a v$ iff i) $w \sim_a v$ and ii) if $b \in N^{\leq n}(w)(a)$, then $b \in \widetilde{B}(w)$ iff $b \in \widetilde{B}(v)$.

Taking chances. Prediction update has been defined to allow agents to take all their available information into account in their decision making. In acting upon it, agents act *conservatively*, as the information-dependent adoption policies defined require *absolute certainty*: agents will adopt only when they *know* that enough of the others (will) have adopted.

An alternative to such conservative behavior is a risky one, where agents adopt whenever they *consider it possible* that enough people (will) have adopted. In the revolution example, this means that agents would join the revolution whenever they see a chance that enough of their neighbors (or of the general population) would join.

Such chance taking behavior is captured as follows:

7.6.2. Definition. [Risky Prediction Update]

Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be a sight n finite model, and let $(F, \leq)$ be as in Def. 7.4.5. The *risky prediction update* of $\mathcal{M}$ results in the model $\mathcal{M}' = (\mathcal{W}, \mathcal{A}, N, \widetilde{B}, \theta, \{\sim'_a\}_{a\in\mathcal{A}})$ where:

- $\widetilde{B}$ is such that:
  - $\forall w \in \mathcal{W}, \widetilde{B}(w) = B(w) \cup \{a \in \mathcal{A} : \exists v \sim_a w, \frac{|N(v)(a)\cap\widetilde{B}(v)|}{|N(v)(a)|} \geq \theta\}$
  - $\forall f \in F$, if $\forall w \in \mathcal{W}, f(w) = B(w) \cup \{a \in \mathcal{A} : \exists v \sim_a w, \frac{|N(v)(a)\cap f(v)|}{|N(v)(a)|} \geq \theta\}$, then $\widetilde{B} \leq f$.

  and
- $w \sim'_a v$ iff i) $w \sim_a v$ and ii) if $b \in N^{\leq n}(w)(a)$, then $b \in \widetilde{B}(w)$ iff $b \in \widetilde{B}(v)$.

To suitably capture e.g. a population of "risky revolutionaries", the risky prediction update should be suitably "globalized" by replacing $N(v)(a)$ with $\mathcal{A}$ everywhere in the definition.

Betting that just any uneliminated possibility is in fact the case is very risky behavior. A natural way to weaken the epistemic requirement of absolute certainty while still allowing for uncertainty to exist is to augment our framework with *beliefs*. Modeling beliefs using the *plausibility orders* of [27], a middle ground between conservative and risky prediction update could be defined. The natural definition would make agents adopt if enough neighbors (are predicted to) have adopted *in each of the worlds the agent considers most plausible*, i.e, if the agent believes enough neighbors (are predicted to) have adopted.

Trendsetters versus followers. An assumption build into threshold models in general is that agents are *followers*: even when they anticipate others' behavior with the prediction update, they only "anticipate their future following of others". Agents are thus *reacting* to others' behavior, even when they are reacting fast.

An interesting alternative would be to utilize agents' information to make them proactive instead; to have *trendsetters* instead of followers. Adding a few trendsetters to a network might induce behavior change towards B even when no-one has adopted initially.

A simple trendsetting adoption policy would state that an agent should adopt whenever she knows that *if* she were to adopt, then enough of her neighbors *will adopt afterwards.* Such an adoption policy involves both counterfactual and temporal reasoning, which complicates a predictive version. A non-predictive version may be defined as follows:

7.6.3. Definition. [(a, w)-Counterfactual Behavior]
Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be an ETM with $w \in \mathcal{W}$. Let the (a, w)-counterfactual behavior of $\mathcal{M}$ be

$$B_{C(a,w)}(v) = \begin{cases} B(v) \cup \{a\} & \text{if } v \sim_a w \\ B(v) & \text{else} \end{cases}$$

7.6.4. Definition. [Trendsetter Update] Let $\mathcal{M} = (\mathcal{W}, \mathcal{A}, N, B, \theta, \{\sim_a\}_{a\in\mathcal{A}})$ be an ETM and let $\{\mathcal{F}, \mathcal{T}\}$ be a partition of $\mathcal{A}$ into sets of followers and trendsetters. The *trendsetter update* of $\mathcal{M}$ *is the ETM* $\mathcal{M}' = (\mathcal{W}, \mathcal{A}, N, B', \theta, \{\sim_a\}_{a\in\mathcal{A}})$ with B' given by $\forall w \in \mathcal{W}$

$$\begin{aligned} B'(w) = &B \cup \left\{a \in \mathcal{F} : \forall v \sim_a w, \frac{|N(v)(a) \cap B|}{|N(v)(a)|} \geq \theta\right\} \cup \\ &\left\{a \in \mathcal{T} : \forall v \sim_a w, \frac{|N(v)(a) \cap B_{C(a,v)}(v)'|}{|N(v)(a)|} \geq \theta\right\} \end{aligned}$$

where $B_{C(a,v)}(v)'$ is the (a, v)-*counterfactual behavior set of* $\mathcal{M}$ *after informed update.*

The trendsetter update may of course also be defined in global and risky versions.

7.7 Conclusion

7.7.1 Summary

The chapter has focused on two intertwined objectives. On the one hand, we have developed logical frameworks for diffusion dynamics of the behavior of agents in social networks, and on the other hand we have developed models for the diffusion dynamics under uncertainty. We focussed our attention on agents which increasing cognitive abilities. At the beginning of this chapter, our threshold models did only focus on the adopting behavior of agents while in the following sections we have equipped agents with epistemic power and also predictive epistemic powers. In the following paragraphs, we summarize our findings.

Threshold Models. The static setting of threshold models may already be described adequately using a propositional logic with proposition symbols that are indexed by agents. On finite networks, threshold ratios may be encoded together with other important structural notions, such as clusters of particular density. As the dynamics of threshold model update is deterministic and state dependent, these may be described using a dynamic modality reducible to the static language. The dynamic modality therefore does not add any expressive power, though it does add explicitness and convenience. We have shown that the logic for threshold-limited influence is sound and complete, and as the static fragment is stated in simple propositional logic, one sees that this logic is also decidable.

Epistemic Threshold Models. Given the propositional logical representation of networks, the epistemic extension of the logic for threshold-limited influence works as expected. As both the diffusion and learning mechanism in the informed update are deterministic and state dependent, the dynamic process that is induced by the dynamic operation can be captured by a reducible dynamic modality. We have shown the epistemic logic of threshold-limited influence to be both sound and complete.

In epistemic threshold models, if agents' behavior is dictated by that of their direct neighbors, then knowledge of more distant agents is redundant. To act under standard threshold model dynamics, however, knowledge of neighbors' behavior is required. If this information is not available, the diffusion speed decreases. In the limit case where no information is available, the diffusion stops. Taken together, the most economical epistemic interpretation of standard threshold models is that their dynamics embodies an *implicit* epistemic assumption that the network structure and behavior of agents at distance 1 is known.

Epistemic Threshold Models with Prediction Update. Prediction update allows agents to fully utilize their information in deciding if and when to adopt a spreading behavior. Describing the dynamics of prediction update requires a dynamic fixed point operator, which is atypical of dynamic epistemic logic. Here we have shown that formulas including this operator are not reducible to the static language. The dynamic operator which is studied in the context of our prediction update thus strictly adds expressive power. The learning mechanisms of prediction update and informed update are identical, but given the fixed point construction involved in the former, obtaining a complete logic is a complex task and is left for future research. We have stated a fixed point axiom and a least fixed point inference rule which were shown to be sound.

7.7.2 Further Research

In future research we plan to work on a full comparative analysis of the different update processes that we have outlined in this chapter. While convergence can be obtained for all different dynamic processes, among the ones we studied, the prediction dynamics will be the fastest in its convergence. In the limit case, where the network and behavior distribution is common knowledge, the prediction update jumps in one step to the fixed point of the standard threshold model update.

The logical treatment of threshold models and their epistemic extension undertaken also yields several more options for further development. Beyond the open problem about a complete logic for prediction update, we see three main directions for further research: A) The logical apparatus and the epistemic extension of the possible generalizations of threshold models discussed in Subsection 7.2.4 are yet to be developed. B) The alternative diffusion processes introduced in Section 7.6 are to be further explored, both on the logical and on the set theoretic level. Their logics may be developed, and their dynamics may be investigated with respect to limit behavior and speed of possible stabilization. C) The epistemic and predictive treatment of non-increasing behaviors is yet to be investigated. Allowing agents to freely *unadopt* radically changes the limit behavior of systems by introducing the possibility of cyclic dynamics, as we have encountered in the previous chapters. Understanding the epistemics of such oscillating limit behavior requires tools going beyond the fixed point oriented mathematics of the current work.

Conclusion to Part III

In this part of the thesis, we have abstracted away from specific case studies of social phenomena of opinion change. We have designed logical systems capturing a wider classes of diffusion phenomena and how they interact with the knowledge of agents in the network. In particular, we have proposed a general complete dynamic hybrid logic to reason about the dynamics of locally driven diffusion phenomena, with the advantage of allowing us to plug-in various dynamic transformations, corresponding to different update rules of diffusion.

We have then extended this framework with a dynamic epistemic dimension, to capture the specific type of learning induced by diffusion dynamics induce learning. We have applied this enriched logic to the cases of opinion change under social influence which had been discussed earlier. We have shown how agents designed to be opaque to each other in the previous part of the thesis might actually come to see-through each other, as a result of social influence dynamics. To reconcile ontic and epistemic dimension, we have also introduced a first knowledge-dependent notion of influence.

Finally, after capturing how diffusion induces learning, we have turned to the converse question of how knowledge affects diffusion. We have designed a propositional dynamic logic to capture diffusion in standard threshold models. We have shown how this simpler logic is able to capture some properties of networks, such as the existence of clusters of a given density, and their relationship to the diffusion behavior. By extending this minimal setting with epistemic the link between the existence dynamic epistemic logic for diffusion in standard threshold-models. We have shown how an increase of information about the network structure and about the state of other agents may accelerate diffusion, assuming more sophisticated agents than in the previous chapters, who can anticipate the behavior of others. Hence, in this part of the thesis, we have made our logics more general, and our agents smarter.

In part IV, we will give a summary of the whole thesis together with a few directions for future research.

Part IV

Perspectives

Chapter 8

Conclusion and Outlook

8.1 High-Level Summary of Results

The work presented in this thesis is a first step toward a general logical understanding of networks dynamics. The order of the chapters takes the reader along the path of our investigation, including its detours. We have taken pains to structure the thesis text in a way that made its topic flow clear at various stages in chapters and parts. Therefore, let us just give a brief high-level summary of our main findings at the end of this road.

We started by focusing on specific social phenomena where the fact that agents influence each other results in a negative collective outcome. Our examples of informational cascades and pluralistic ignorance focused on the spread of opinions, driven by observed incoming information. We found general features, for various notions of rationality, that enabled us to prove that perfectly rational individuals can lead a whole group to make the wrong choice.

We then showed how dynamic properties of social phenomena, such as their degree of fragility, usually depend on the way the influence network is structured. Moreover, we found a precise sense in which collective failure phenomena rely on the typical "intermediate" situation of agents in social contexts: they share too much information not to influence each other, but not enough information to fully actualize their potential group knowledge. We concluded that a general logic for network dynamics should be able to talk both about social network structure and about what information agents receive about each other's behavior.

Abstracting away from specific case studies of opinion change, we then moved toward the design of suitable logics that can capture wider classes of diffusion phenomena over social networks, as well as their interaction with information. In particular, we designed a complete dynamic hybrid logic that can reason about locally driven dynamic phenomena in a general fashion, allowing to "plug-in" various transformations rules, corresponding to different diffusion dynamics.

Next, we considered how diffusion dynamics affect what agents know about each other, i.e. how diffusion processes induce learning. To capture the relevant ontic and epistemic changes, we extended our hybrid framework with tools from dynamic-epistemic logic. The resulting richer logic of social learning processes was illustrated by returning to the earlier cases of opinion change under social influence. We found that agents assumed to be "opaque" to each other might actually come to "see-through" each other. More precisely, by interpreting changes of behavior of their neighbors and knowing the rules of social influence, agents can sometimes come to know more about each other than what they can initially observe. This finding exemplified how diffusion may induce an increase of information.

Finally, we also investigated the interaction between diffusion phenomena and knowledge from a different angle. We studied the converse issue of how knowledge affects diffusion, and in particular we were able to prove formal results that show how more information may accelerate diffusion. In particular, we were able to show how agents in threshold models with unbounded reasoning abilities can predict, and anticipate, the spread of social behavior.

Along the way, we played with different levels of generality, different formal tools, short term and long term dynamics, and different assumptions concerning influence, agency, and rationality. While we have not settled on one unique framework, something that may not even be a good thing to pursue, we hope to have illustrated how logical methods can capture laws of networks dynamics, at levels of generality that naturally mirror various grain levels for describing social phenomena in different languages.

Thus, step by step, we hope to have shown how logic brings fresh perspectives, and perhaps even a measure of abstract ordering, to the vast world of dynamic processes over social networks.

Clearly though, we are just at the start, and we have not presented a logical theory that matches the richness of social reality. In our view, there are at least two things lacking for such a theory that lie beyond the work presented here. One is further empirical coverage of all relevant aspects of social agency, the other is more abstract mathematical foundations behind the various logics that we have applied in our case studies. In the second section of this final chapter we give some thoughts on both, referring to ongoing research of the author that was not yet in a state to warrant inclusion in this dissertation.

8.2 Further Directions

Our ongoing research follows two main directions. The first is about technical foundations, searching for a logical theory of *dynamical systems*, the mathematics in the background of much social network theory. What are fundamental qualitative laws in this area, and what logical languages are best suited to capture these? The second direction is about extended coverage: agents in networks do not just process information and form opinions, they also pursue goals. Hence it is entirely natural to merge the perspective of this thesis with that of *game theory*, where goals and strategic behavior come into their own. In this section, we very briefly present some of our current observations on both directions.[1]

8.2.1 Stability, Oscillations, and Graph Properties

Generally speaking, all phenomena studied in this thesis are about how a given static network model – a distribution of properties of nodes on a graph – evolves through time, under application of some uniform deterministic local update rule. The rule determines a node's state at the next moment, given the current state of the accessible nodes around it. The result of applying this rule once on all nodes at the same time in parallel is a unique new model; and the result of iterating this application is a unique sequence of models. In the long run, such a sequence may stabilize to a unique model, or it may not. If not, it may enter an infinite loop, or exhibit oscillations.[2]

Our ongoing work [61] concerns long term behavior of classes of update rules. What are conditions on the format of update rules that lead to stabilization, or make this fail – and given that the social network structure tends to be equally important for this, what interplay of logical rule and network structure determines limit behavior? We merely list some points in our current approach.

Our current approach to the format of update rules is to treat them as "modal automata", an example of which is found in the opinion change rule of [145] introduced in Section 2.2.2. Many of the usual update rules that have been studied in the logical literature on social networks are definable in systems of modal logic, sometimes graded modal logic (that can count numbers of accessible points with some property), and only rarely richer hybrid, first-order, or higher-order languages. Now we can show that basic modal formulas correspond to "semantic automata" (using a suitable modification of the approach in [40] that check for truth of the formula in finite pointed models. One can think of such automata as highly bounded agents, and stepping up in the hierarchy of logical definability means ascending up to more powerful automata.

As for describing the long-term dynamic behavior produced by these automata that work stepwise, we are currently exploring two frameworks: dynamic-epistemic logics of substitutions in the style of [47] enriched with an iteration operator, and extensions of modal fixed-point logics such as the μ-calculus with

[1] This work is documented in two working papers [61, 59]. Part of this project is joint work with Johan van Benthem, whose [46] provides a technical program for studying both dynamical systems and social network games in a setting of dynamic-epistemic, temporal, and fixed-point logics.

[2] On finite network graphs, it is always the case that the sequence either stabilizes or oscillates, as there are only finitely many models to cycle through. On infinite networks, there is a third possibility of "divergence".

an "oscillation operator" (cf. [46]).[3][4]

However, we are also exploring an opposite angle, focusing on limit behavior as arising from network structure rather than logical form of the update rule. Just as an illustration, fix the "Unanimity Rule" where an agent adopts property p only if all its neighbors have p right now, and drops p only if none of his neighbors has p now. We can show that a population behavior oscillates under this rule if and only if agents with p and agents without p are distributed as a proper 2-coloring of the network graph, that is, when a model validates the modal formula. Hence, oscillating behavior can only occur over graphs which are properly 2-colorable. In other words, there is a class of graphs which guarantees stabilization: the class of non-2-colorable graphs, i.e any graph which contains a cycle of odd-length. Thus, we also see how graph theoretic notions and facts are involved in the study of the topics in this thesis.

We can even go one step further here, and bring in logic at a new abstraction level, more in the sense of foundations of mathematics. Can colorability be defined in the logical languages that we have considered for network dynamics, and can its basic theory be axiomatized completely? While colorability is not definable in basic modal logic (or in any bisimulation invariant extension of it), it is easy to show that *non*-k-colorability is definable only relative to the class of graphs with at most n nodes, or by adding the universal modality to the language. We are currently investigating what other tools from logic have to say about colorability, among which hybrid logic[5], PDL, and non-standard inference rules in the style of Gabbay's rule for irreflexivity [79].

We hope to have given at least some flavor of where more foundational logical studies of network dynamics might go. For formal definitions and theorems behind the claims made so far, we refer to the working manuscripts [61, 60].

8.2.2 From Network Diffusion to Network Games

Social networks studied in the preceding dynamical systems mode involve deterministic update rules, leaving agents no choices. Now in this thesis, we did study many scenarios where intuitively, agents do deliberate about a best response to what they know and observe about their environment. Making this feature explicit takes us from automated agents to deliberating agents that have choices an goals, i.e., to the area of *game theory*.[6]

---

[3] Another logical approach to dynamical systems for social networks is found in [110], where the model space of dynamic-epistemic logic gets metrized.

[4] Eventually, we may also want to use the dynamic topological logic of [116] that integrates dynamic update rules with topological structure inside the network.

[5] Here we have been inspired by the work on modal logics for elementary graph theory in [38, 39].

[6] Of course, this same border-line occurs inside game theory, e.g., in the contrast between the rich view of reasoning agents in epistemic foundations of classical game theory [130] versus

An obvious sort of game that fits very well with social network are the *iterated Boolean games* [91, 96]. Here agents can choose their next action based on observing their environment, while they have qualitative goals that come in the form of logical formulas defining properties of histories that the agents consider important (in some hierarchical order that induces a preference eon histories much like in the criterion-based preference of [119]).[7] With this set-up, we can apply the standard notions of game theory, and look at Nash equilibria as a way of measuring whether social agents with choices of update can achieve strategic equilibria satisfying individual (and perhaps also social) goals. Of yet further interest is the use of "finite-automata strategies" in [91] as a way of restricting attention to bounded agents.

We can see this framework as an obvious extension of our earlier setting, with agents that are now free to choose. This makes the concerns in the preceding subsection a special case: the dynamics determined by an update rule corresponds to a unique strategy profile, and the sequence generated by iteratively updating a model with a given rule corresponds to a unique history of the game. We can show, for instance, that for each given update rule (such as the unanimity rule above), making satisfaction of the formula describing stability the goal of all agents leads to stabilization being a Nash equilibrium for the associated Boolean game.

We can also show that finite-automaton strategies in Boolean games for networks are in one-to-one correspondence with local update rules of the sort we discussed earlier. While this may make our earlier framework look like an extreme, and somewhat narrow-minded case of agents chained to automated predetermined responses, we can also see our earlier theory as one of strategy profiles, that arose out of epistemic deliberation of the sort discussed in the main body of this thesis.

Now social networks do suggest one type of crucial structure that is not explicitly present in Boolean games, namely the neighborhood relation that regulates communication and influence.[8] One can define social network games exploiting this structure that seem to go beyond Boolean games.[9] However, [147], which represents work done independently from our ongoing program, shows that natural social network games are equivalent to iterated Boolean games, through a series of clever encodings where network structure can be absorbed into the propositional variables controlled by the players.[10]

---

the sparse views based on evolutionary dynamics favored in evolutionary game theory [103].

[7]Actions in Boolean games amount to assigning truth values to atomic propositions under the player's control. This may seem quite austere, but it does fit many of our network scenarios.

[8]For the importance of this relation in dynamical systems for evolutionary game theory, see [155].

[9]See the survey article [106] for work that has already been done in this area.

[10]Several recent independent lines of work bring together game and networks from different angles. One example is given by [157], which studies "formation network games". Combining

We do not believe that this equivalence is the end of the story, since there are many natural games that can be played over networks, witness the variety of games that is possible already on simple epistemic models. To see this, the reader may consult the various logical games discussed in [45], or the "knowledge games" of [3].

In particular, we believe that game scenarios will enable us to bring the rich literature on logics of information, action, and preference in games to bear on the study of social networks in a more systematic manner than we have done in this thesis.

While we have not presented any technical results in this discussion, the reader may have acquired a sense of a broader world behind social networks. We think of the resulting situation in two complementary ways. One can view game scenarios as richer description level underneath rougher or poorer logics of social phenomena, but one can equally well, in line with our reading of update rules as strategies, view the logics studied in this thesis as well-chosen useful high-level descriptions of patterns of behavior emerging from games. Again, for further detail, we refer to a working manuscript, this time [59].

ideas from game theory, network theory [68] and formal learning theory [107, 98, 81], another example is [82], which models how agents in a network can collectively learn via iterated games.

Bibliography

[1] Andreea Achimescu. Games and logics for informational cascades. Master's thesis, ILLC, 2014. Master of Logic Thesis, MoL-2014-04.

[2] Andreea Achimescu, Alexandru Baltag, and Joshua Sack. The probabilistic logic of communication and change. *Journal of Logic and Computation*, To appear.

[3] Thomas Agotnes and Hans van Ditmarsch. What will they say? public announcement games. *Synthese (KRA)*, 179(Suppl.1):57–85, 2011.

[4] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985.

[5] Lisa R. Anderson and Charles A. Holt. Classroom games: Information cascades. *The Journal of Economic Perspectives*, 10(4):187–193, 1996.

[6] Lisa R. Anderson and Charles A. Holt. Information cascades in the laboratory. *The American Economic Review*, 87(5):847–862, 1997.

[7] Krzysztof R. Apt, Davide Grossi, and Wiebe van der Hoek. Epistemic protocols for distributed gossiping. In R. Ramanumam, editor, *Proceedings of the Conference on Theoretical Aspects of Rationality and Knowledge*, TARK 2015. ACM, 2015.

[8] Krzysztof R. Apt and Evangelos Markakis. Diffusion in Social Networks with Competing Products. In G. Persiano, editor, *SAGT 2011*, LNCS 6982, pages 212–223. Springer, 2011.

[9] Krzysztof R. Apt, Andreas Witzel, and Jonathan A. Zvesper. Common knowledge in interaction structures. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '09, pages 4–13, New York, NY, USA, 2009. ACM.

[10] Carlos Areces and Balder ten Cate. Hybrid logics. In Patrick Blackburn, Johan van Benthem, and Frank Wolter, editors, *Handbook of Modal Logic*, pages 821–868. Elsevier, Amsterdam, 2007.

[11] Maduka Attamah, Hans van Ditmarsch, and Davide Grossi. Knowledge and gossip. In *Proceedings of the 21st ECAI*, pages 21–26. IOS Press, 2014.

[12] Guillaume Aucher, Philippe Balbiani, Luis Fariñas del Cerro, and Andreas Herzig. Global and local graph modifiers. *Electronic Notes in Theoretical Computer Science*, 231:293–307, 2009. Proceedings of the 5th Workshop on Methods for Modalities (M4M5 2007).

[13] Robert J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(5):1236–1239, 1976.

[14] Robert Axelrod. *Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton University Press, Princeton, New Jersey, 1997.

[15] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider. *The Description Logic Handbook – Theory, Implementation and Applications*. Cambridge University Press, UK, second edition, 2007.

[16] Alexandru Baltag. Logics for insecure communication. In Johan van Benthem, editor, *Proceedings of TARK'01 (Eight Conference on Rationality and Knowledge)*, pages 111–122, 2001.

[17] Alexandru Baltag. A logic for suspicious players: Epistemic actions and belief updates in games. *Bulletin of Economic Research*, 54:1–46, 2002.

[18] Alexandru Baltag, Zoé Christoff, Jens Ulrik Hansen, and Sonja Smets. Logical models of informational cascades. In Johan van Benthem and Fenrong Lui, editors, *Logic across the University: Foundations and Applications*, Studies in Logic, pages 405–432. College Publications, 2013.

[19] Alexandru Baltag, Zoé Christoff, Rasmus K. Rendsvig, and Sonja Smets. Dynamic epistemic logic of diffusion and prediction in social networks. Technical Report PP-2015-22, ILLC, Amsterdam, 2015.

[20] Alexandru Baltag, Nina Gierasimczuk, and Sonja Smets. Belief revision as a truth-tracking process. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK XIII, pages 187–190, New York, NY, USA, 2011. ACM.

[21] Alexandru Baltag, Nina Gierasimczuk, and Sonja Smets. Truth-tracking by belief revision. Technical Report PP-2014-20, ILLC, Amsterdam, 2014. Submitted for publication.

[22] Alexandru Baltag, Nina Gierasimczuk, and Sonja Smets. On the solvability of inductive problems: A study in epistemic topology. In R. Ramanumam, editor, *Proceedings of the Conference on Theoretical Aspects of Rationality and Knowledge*, TARK 2015. ACM, 2015.

[23] Alexandru Baltag and Lawrence Moss. Logics for epistemic programs. *Synthese*, 139(2):165–224, 2004.

[24] Alexandru Baltag, Lawrence Moss, and Slawomir Solecki. The logic of public announcements, common knowledge and private suspicions. In *Proceedings of TARK'98 (Seventh Conference on Theoretical Aspects of Rationality and Knowledge)*, pages 43–56. Morgan Kaufmann Publishers, 1998.

[25] Alexandru Baltag, Ben Rodenhäuser, and Sonja Smets. Doxastic attitudes as belief-revision policies. In *Online pre-proceeding of the workshop on Strategies for Learning, Belief Revision and Preference Change*. ESSLLI 2012, 2012.

[26] Alexandru Baltag and Sonja Smets. Dynamic belief revision over multi-agent plausibility models. In M. Woolridge G. Bonanno, W. van der Hoek, editor, *Pre-proceedings of the 7th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 2006)*, 2006.

[27] Alexandru Baltag and Sonja Smets. A Qualitative Theory of Dynamic Interactive Belief Revision. In Giacomo Bonanno, Wiebe van der Hoek, and Michael Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, Texts in Logic and Games, Vol. 3, pages 9–58. Amsterdam University Press, 2008.

[28] Alexandru Baltag and Sonja Smets. The logic of conditional doxastic actions. In Robert van Rooi and Krysztof Apt, editors, *New Perspectives on Games and Interaction*, Texts in Logic and Games 4. Amsterdam University Press, 2008.

[29] Alexandru Baltag and Sonja Smets. Group belief dynamics under iterated revision: Fixed points and cycles of joint upgrades. In *Proceedings of Theoretical Aspects of Rationality and Knowledge TARK*, 2009.

[30] Alexandru Baltag and Sonja Smets. Keep changing your beliefs, aiming for the truth. *Erkenntnis*, 75:255–270, 2011.

[31] Alexandru Baltag and Sonja Smets. Protocols for belief merge: Reaching agreement via communication. *Logic Journal of IGPL*, 21(3):468–487, 2013.

[32] Alexandru Baltag and Sonja Smets, editors. *Johan van Benthem on Logic and Information Dynamics*, volume 5 of *Outstanding Contributions to Logic Series*. Springer, 2014.

[33] Alexandru Baltag, Hans van Ditmarsch, and Lawrence Moss. Epistemic logic and information update. In Pieter Adriaans and Johan van Benthem, editors, *Handbook on the Philosophy of Information*. Elsevier, 2008.

[34] Abhijit V. Banerjee. A simple model og herd behavior. *The Quarterly Journal of Economics*, 107(3):797—817, 1992.

[35] Abhijit V. Banerjee, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. The diffusion of microfinance. *Science*, 341(6144), 2013.

[36] Jon Barwise. Three views of common knowledge. In *Proceedings of the 2nd conference on Theoretical aspects of reasoning about knowledge*, TARK '88, pages 365–379, San Francisco, CA, USA, 1988. Morgan Kaufmann Publishers Inc.

[37] Pierpaolo Battigalli, Adam Brandenburger, Amanda Y. Friedenberg, and Marciano Siniscalchi. *Epistemic Game Theory: Reasoning about Strategic Uncertainty*. In preparation.

[38] Mario R. F. Benevides and L. Menasché Schechter. Modal expressiveness of graph properties. *Electron. Notes Theor. Comput. Sci.*, 205:31–47, April 2008.

[39] Mario R. F. Benevides and L. Menasché Schechter. Using modal logics to express and check global graph properties. *Logic Jnl IGPL*, 17 (5):559–587, 2009.

[40] Johan van Benthem. *Essays in Logical Semantics*. Number Vol. 29 in Studies in Linguistics and Philosophy. D. Reidel Publishing Company, 1986.

[41] Johan van Benthem. One is a lonely number: on the logic of communication. In Z. Chatzidakis, P. Koepke, and W. Pohlers, editors, *Proceedings of the Amsterdam Colloquium '02*, pages 96–129, 2006.

[42] Johan van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007.

[43] Johan van Benthem. Rational dynamics and epistemic logic in games. *International Game Theory Review*, 9(1):13–45, 2007. Erratum reprint, Volume 9:2, 377–409.

[44] Johan van Benthem. *Logical Dynamics of Information and Interaction.* Cambridge University Press, 2011.

[45] Johan van Benthem. *Logic in Games.* MIT Press, 2014.

[46] Johan van Benthem. Oscillations, logic, and dynamical systems. In Sujata Ghosh and Jakub Szymanik, editors, *The Facts Matter. Essays on Logic and Cognition in Honour of Rineke Verbrugge*, pages 9–22. College Publications, 2015.

[47] Johan van Benthem, Jan van Eijck, and Barteld Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, 2006.

[48] Johan van Benthem, Jelle Gerbrandy, and Barteld Kooi. Dynamic update with probabilities. *Studia Logica*, 93(1):67–96, 2009.

[49] Johan van Benthem and Fenrong Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-classical Logics*, 17(2):157–182, 2007.

[50] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.

[51] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3):151–170, 1998.

[52] Jens Christian Bjerring, Jens Ulrik Hansen, and Nikolaj Jang Lee Linding Pedersen. On the rationality of pluralistic ignorance. *Synthese*, 191:2445–2470, 2014.

[53] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic.* Cambridge Tracts in Theoretical Computer Science, 53. Cambridge University Press, United Kingdom, 2001.

[54] Patrick Blackburn and Balder ten Cate. Pure extensions, proof rules, and hybrid axiomatics. *Studia Logica*, 84(2):277–322, 2006.

[55] Thomas Bolander and Mikkel Birkegaard Andersen. Epistemic planning for single- and multi-agent systems. *Journal of Applied Non-Classical Logics*, 21(1):9–34, 2011.

[56] Richard Bradley. Reaching a consensus. *Social Choice and Welfare*, 29(4):609–632, 2007.

[57] Torben Braüner. Hybrid logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2008 edition, 2008.

[58] Zoé Christoff. A logic for social influence through communication. In Emiliano Lorini, editor, *Proceedings of Eleventh European Workshop on Multi-Agent Systems (EUMAS 2013)*, volume 1113, pages 31–39. CEUR, 2013.

[59] Zoé Christoff. Boolean games and social influence, manuscript.

[60] Zoé Christoff. A note on modal definability of colorability. manuscript.

[61] Zoé Christoff. Update rules and oscillations in networks, manuscript.

[62] Zoé Christoff and Jens Ulrik Hansen. A two-tiered formalization of social influence. In Huaxin Huang, Davide Grossi, and Olivier Roy, editors, *Logic, Rationality and Interaction, Proceedings of the Fourth International Workshop (LORI 2013)*, volume 8196 of *Lecture Notes in Computer Science*, pages 68–81. Springer, 2013.

[63] Zoé Christoff and Jens Ulrik Hansen. A logic for diffusion in social networks. *Journal of Applied Logic*, 13(1):48–77, 2015.

[64] Zoé Christoff, Jens Ulrik Hansen, and Carlo Proietti. Reflecting on social influence in networks. *Journal of Logic, Language and Information*, 2016. To appear.

[65] Giovanni Ciná and Ulle Endriss. A syntactic proof of arrow's theorem in a modal logic of social choice functions. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2015)*, 2015.

[66] Marie Jean Antoine Nicolas de Caritat, Marquis de Condorcet. *Essai sur l'application de l'analyse la probabilité des décisions rendues à la pluralité des voix*. Paris: L'Imprimerie Royale, 1785.

[67] Brian A. Davey and Hilary A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, second edition, 2002.

[68] Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[69] Hans Ditmarsch, Jan van Eijck, Pere Pardo, Rahim Ramezanian, and François Schwarzentruber. Gossip in dynamic networks. Under Submission, 2015.

[70] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Syntese Library volume 337. Springer, The Netherlands, 2008.

[71] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, USA, 2010.

[72] Jan van Eijck. Gossip in dynamic networks, 2015. Invited talk at the 3rd Tsinghua Logic Colloquium, Beijing, October 21 2015. http://www.jaist.ac.jp/ v-sano/jw2014/index.html.

[73] Jan van Eijck and Floor Sietsma. Message-generated kripke semantics. In *Proceedings of the 10th International Conference on Autonomous agents and Multiagent Systems (AAMAS 2011)*, pages 1183–1184, 2011.

[74] Ronald Fagin and Joseph Y. Halpern. Modelling knolwedge and action in distributed systems. *Distributed Computing*, 3(4):159–177, 1989.

[75] Ronald Fagin and Joseph Y. Halpern. Reasoning about knowledge and probability. *Journal of the ACM*, 41(2):340–367, 1993.

[76] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about knowledge*. MIT Press, Cambridge, MA, USA, 1995.

[77] James M. Fields and Howard Schuman. Public beliefs about the beliefs of the public. *The Public Opinion Quarterly*, 40(4):427–448, 1976.

[78] Kit Fine. In so many possible worlds. *Notre Dame Journal of Formal Logic*, 13(4):516–520, October 1972.

[79] Dov M. Gabbay. An irreflexivity lemma with applications to axiomatizations of conditions on tense frames. In Uwe Mnnich, editor, *Aspects of Philosophical Logic*, volume 147 of *Synthese Library*, pages 67–89. Springer Netherlands, 1981.

[80] James Gate and Iain A. Stewart. The expressibility of fragments of hybrid graph logic on finite digraphs. *Journal of Applied Logic*, 11(3):272–288, 2013.

[81] Nina Gierasimczuk. *Knowing one's limits: Logical Analysis of Inductive Inference*. PhD thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands, 2010. ILLC Dissertation Series DS-2010-11.

[82] Nina Gierasimczuk and Sanne Kosterman. Collective learning in games through social networks. In *Proceedings of the 1st International Workshop on Social Influence Analysis, SocInf 2015, co-located with 24th International Joint Conference on Artificial Intelligence (IJCAI 2015), Buenos Aires, Argentina, July 27, 2015*, pages 35–41, 2015.

[83] Patrick Girard, Jeremy Seligman, and Fenrong Liu. General dynamic dynamic logic. In Thomas Bolander, Torben Brauner, Silvio Ghilardi, and Lawrence Moss, editors, *Advances in Modal Logic, Volume 9*, pages 239–260. College Publication, 2012.

[84] Alvin I. Goldman. *Knowledge in a Social World.* Oxford Univ Press, 1999.

[85] Alvin I. Goldman. A guide to social epistemology. In Alvin I. Goldman and Dennis Whitcomb, editors, *Social Epistemology – Essential Readings*, pages 11–37. Oxford University Press, New York, 2011.

[86] Alvin I. Goldman and Dennis Whitcomb, editors. *Social Epistemology, Essential Readings.* Oxford University Press, 2011.

[87] Sanjeev Goyal. Interaction Structure and Social Change. *Journal of Institutional and Theoretical Economics*, 152(3):472–494, 1996.

[88] Umberto Grandi and Ulle Endriss. First-order logic formalisation of impossibility theorems in preference aggregation. *Journal of Philosophical Logic*, 42(4):595–618, 2013.

[89] Umberto Grandi, Emiliano Lorini, and Laurent Perrussel. Propositional opinion diffusion. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '15, pages 989–997, Richland, SC, 2015. International Foundation for Autonomous Agents and Multiagent Systems.

[90] Mark Granovetter. Threshold Models of Collective Behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.

[91] Julian Gutierrez, Paul Harrenstein, and Michael Wooldridge. Iterated boolean games. *Inf. Comput.*, 242(C):53–79, June 2015.

[92] Jonathon R. B. Halbesleben and M. Ronald Buckley. Pluralistic ignorance: historical development and organizational applications. *Management Decision*, 42(1):126–138, 2004.

[93] Jonathon R. B. Halbesleben, Anthony R. Wheeler, and M. Ronald Buckley. Understanding pluralistic ignorance: application and theory. *Journal of Managerial Psychology*, 22(1):65–83, 2007.

[94] Jens Ulrik Hansen. Reasoning about opinion dynamics in social networks. In *Proceedings of the eleventh conference on logic and the foundations of game and decision theory (LOFT XI)*. Bergen, Norway, July 27–30, 2014.

[95] David Harel, Dexter Kozen, and Jerzy Tiuryn. *Dynamic Logic*. The MIT Press, Cambridge, 2000.

[96] Paul Harrenstein, Wiebe van der Hoek, John-Jules Meyer, and Cees Witteveen. Boolean games. In *Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '01, pages 287–298, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[97] Vincent Hendricks and John Symons. Epistemic logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2009 edition, 2009.

[98] Vincent F. Hendricks. *The Convergence of Scientific Knowledge a View From the Limit*. 2001.

[99] Vincent F. Hendricks, Pelle G. Hansen, and Rasmus K. Rendsvig. Infostorms. *Metaphilosophy*, 44:301–326, 2013.

[100] Vincent F. Hendricks and Pelle Guldborg Hansen. *Infostorms: How to take Information Punches and Save Democracy*. Copernicus Books, Springer, 2014.

[101] Henry Hexmoor. *Computational Network Science, An Algorithmic Approach*. Elsevier, Amsterdam, 2015.

[102] Jaakko Hintikka. *Knowledge and Belief –An Introduction to the Logic of the Two Notions*. Cornell University, 1962.

[103] Joseph Hofbauer and Karl Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[104] Wesley H. Holliday. Dynamic testimonial logic. In *Logic, Rationality, and Interaction, Second International Workshop, LORI 2009, Chongqing, China, October 8-11, 2009. Proceedings*, pages 161–179, 2009.

[105] Matthew O. Jackson. *Social and Economic Networks*. Princeton University Press, 2010.

[106] Matthew O. Jackson and Yves Zenou. Games on networks. In Peyton Young and Shmuel Zamir, editors, *Handbook of Game Theory*, volume 4. Elsevier Science, 2014. Available at SSRN: http://ssrn.com/abstract=2136179.

[107] Kevin Kelly. *The Logic of Reliable Inquiry*. Oxford University Press, 1996.

[108] Kevin Kelly. A computational learning semantics for inductive empirical knowledge. In Alexandru Baltag and Sonja Smets, editors, *Johan van Benthem on Logic and Information Dynamics*, volume 5 of *Outstanding Contributions to Logic Series*, pages 289–337. Springer, 2014.

[109] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *KDD 03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, New York, NY, USA, 2003.

[110] Dominik Klein and Rasmus K. Rendsvig. Model transformers for dynamical systems of dynamic epistemic logic. In Wiebe van der Hoek, Wesley H. Holliday, and Wen fang Wang, editors, *Logid, Rationality and Interaction, Proceedings of the 5th International Workshop, LORI 2015, Taipei, Taiwan, October 28-30, 2015*, volume 9394 of *LNCS*. Springer, 2015.

[111] Phokion Kolaitis. Infinitary Logic in Finite Model Theory. In M. Dalla Chiara, K. Doets, D. Mundici, and J. van Benthem, editors, *Logic and Scientific Methods*, pages 9–58. Springer, 1997.

[112] Barteld Kooi. Probabilistic dynamic epistemic logic. *J. of Logic, Lang. and Inf.*, 12(4):381–408, 2003.

[113] Barteld Kooi and Bryan Renne. Arrow update logic. *The Review of Symbolic Logic*, 4:536–559, 12 2011.

[114] Marcus Kracht, John-Jules Meyer, and Krister Segerberg. The logic of action. In Edward N. Zalta, editor, *Stanford Encyclopedia of Philosophy.* 2008.

[115] David Krech and Richard S. Crutchfield. *Theories and Problems of Social Psychology.* New York: McGraw-Hill, 1948.

[116] Philip Kremer and Grigori Mints. Dynamic topological logic. In Marco Aiello, Ian Pratt-Hartmann, and Johan Van Benthem, editors, *Handbook of Spatial Logics*, pages 565–606. Springer Netherlands, 2007.

[117] Keith Lehrer and Carl Wagner. *Rational Consensus in Science and Society. A Philosophical and Mathematical Study.* D. Reidel Publishing Company, Dordrecht, Holland, 1981.

[118] Fa-Hsien Li, Cheng-Te Li, and Man-Kwan Shan. Labeled Influence Maximization in Social Networks for Target Marketing. *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, pages 560–563, October 2011.

[119] Fenrong Liu. *Reasoning about preference dynamics*, volume 354 of *Synthese Library*. Springer, 2011.

[120] Fenrong Liu, Jeremy Seligman, and Patrick Girard. Logical dynamics of belief change in the community. *Synthese*, 191(11):2403–2431, 2014.

[121] Cheng Long and Raymond Chi-Wing Wong. Minimizing Seed Set for Viral Marketing Paper. *2011 IEEE 11th International Conference on Data Mining*, pages 427–436, December 2011.

[122] Winter A. Mason, Frederica Conrey, and Eliot R. Smith. Situating social influence processes: Multidirectional flows of influence within social networks. *Personality and Social Psychology Review*, 11:279–300, 2007.

[123] John-Jules Ch. Meyer and Wiebe van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge Tracts in Theoretical Computer Science 41. Cambridge University Press, Cambridge, UK, 1995.

[124] Dale T. Miller and Cathy McFarland. Pluralistic ignorance; when similarity is interpreted as dissimilarity. *Journal of Personality and Social Psychology*, 53:298–305, 1987.

[125] Dale T. Miller and Cathy McFarland. When social comparison goes awry: The case of pluralistic ignorance. In J. Suls and T. Wills, editors, *Social comparison: Contemporary theory and research*, pages 287–313. Erlbaum, Hillsdale, NJ, 1991.

[126] Stephen Morris. Contagion. *Review of Economic Studies*, 67:57–78, 2000.

[127] Mark E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[128] Hubert J. O'Gorman. The discovery of pluralistic ignorance: An ironic lesson. *Journal of the History of the Behavioral Sciences*, 22(October):333–347, 1986.

[129] Hubert J. O'Gorman and Stephen L. Garry. Pluralistic ignorance – a replication and extension. *The Public Opinion Quarterly*, 40(4):449–458, 1976.

[130] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge, Massachusetts, 1994.

[131] Eric Pacuit and Samer Salame. Majority logic. In Didier Dubois, Christopher Welty, and Mary-Anne Williams, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Ninth International Conference (KR2004)*, pages 598–605. AAAI Press, 2004.

[132] Andres Perea. *Game Theory: Reasoning and Choice*. Cambridge University Press, 2012.

[133] Jan Plaza. Logics of public communications. In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216. Oak Ridge National Laboratory, 1989.

[134] Deborah A. Prentice and Dale T. Miller. Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *Journal of Personality and Social Psychology*, 64(2):243–256, 1993.

[135] Ben Rodenhäuser. *A Matter of Trust: Dynamic Attitudes in Epistemic Logic*. PhD thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands, 2014. ILLC Dissertation Series DS-2014-04.

[136] Floris Roelofsen. Exploring logical perspectives on distributed information and its dynamics. Master's thesis, ILLC, University of Amsterdam, 2005.

[137] Hans Rott. Conditionals and theory change: revisions, expansions, and additions. *Synthese*, 81(1):91–113, 1989.

[138] Olivier Roy. *Thinking Before Acting. Intentions, Logic, Rational Choice.* PhD thesis, Institute for logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands, February 2008. ILLC Dissertation series DS-2008-03.

[139] Ji Ruan and Michael Thielscher. A logic for knowledge flow in social networks. In Dianhui Wang and Mark Reynolds, editors, *AI 2011: Advances in Artificial Intelligence*, volume 7106 of *Lecture Notes in Computer Science*, pages 511–520. Springer Berlin Heidelberg, 2011.

[140] Bryce Ryan and Neal C. Gross. The diffusion of hybrid seed corn in two Iowa communities. *Rural sociology*, 8(1):15–24, 1943.

[141] Katsuhiko Sano and Satoshi Tojo. Dynamic epistemic logic for channel-based agent communication. In Kamal Lodaya, editor, *Logic and Its Applications*, volume 7750 of *Lecture Notes in Computer Science*, pages 109–120. Springer Berlin Heidelberg, 2013.

[142] Thomas C. Schelling. Models of segregation. *The American Economic Review*, 59(2):488–493, 1969.

[143] Thomas C. Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186, 1971.

[144] Thomas C. Schelling. *Micromotives and Macrobehavior*. Norton, New York, 1978.

[145] Jeremy Seligman, Fenrong Liu, and Patrick Girard. Logic in the community. In Mohua Banerjee and Anil Seth, editors, *Logic and Its Applications*, volume 6521 of *Lecture Notes in Computer Science*, pages 178–188. Springer, 2011.

[146] Jeremy Seligman, Fenrong Liu, and Patrick Girard. Facebook and the epistemic logic of friendship. In Burkhard C. Schipper, editor, *Proceedings of the 14th Conference on Theoretical Aspects of Reasoning about Knowledge*, TARK 2013, pages 207–222, 2013.

[147] Jeremy Seligman and Declan Thompson. Boolean network games and iterated boolean games. In Wiebe van der Hoek, Wesley H. Holliday, and Wen-fang Wang, editors, *Logic, Rationality, and Interaction*, volume 9394 of *Lecture Notes in Computer Science*, pages 353–365. Springer Berlin Heidelberg, 2015.

[148] Paulo Shakarian, Sean Eyre, and Damon Paulo. A Scalable Heuristic for Viral Marketing Under the Tipping Model. 2013.

[149] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems:Algorithmic, Game Theoretic and Logical Foundations.* Cambridge University Press, Cambridge, UK, 2009.

[150] Floor Sietsma. *Logics of Communication and Knowledge*. PhD thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands, 2014. ILLC Dissertation Series DS-2012-11.

[151] Floor Sietsma and Jan van Eijck. Message passing in a dynamic epistemic logic setting. In Krzysztof Apt, editor, *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK 2011, pages 212–220, 2011.

[152] Pramesh Singh, Sameet Sreenivason, Boleslaw K. Szymanski, and Gyorgy Korniss. Threshold-limited spreading in social networks with multiple initiators. *Scientific Reports*, 3(2330), 2013.

[153] Brian Skyrms. *The Dynamics of Rational Deliberation.* Oxford University Press, 1990.

[154] Brian Skyrms. *Signals: Evolution, Learning & Information.* Oxford University Press, Forthcoming.

[155] Brian Skyrms and Robin Pemantle. A dynamic model of social network formation. *Proceedings of the National Academy of Sciences*, 97(16):9340–9346, 2000.

[156] Paul Smolensky and Géraldine Legendre. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. MIT Press, 2011.

[157] Martijn Tennekes. *Network Formation Games*. PhD thesis, Universiteit Maastricht, Maastricht, The Netherlands, 2010.

[158] Peter Vanderschraaf and Giacomo Sillari. Common knowledge. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2007 edition, 2007.

[159] Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.

[160] Yanjing Wang and Qinxiang Cao. On axiomatizations of public announcement logic. *Synthese*, 190(1):103–134, 2013.

[161] Yanjing Wang, Floor Sietsma, and Jan van Eijck. Logic of information flow on communication channels. In *Logic and Interactive Rationality, Yearbook 2009*, Amsterdam, 2010. ILLC.

[162] Lara van Weegen. Informational cascades under variable reliability assessments. a formal and empirical investigation. Master's thesis, ILLC, 2014. Master of Logic Thesis, MoL-2014-21.

[163] James D. Westphal and Michael K. Bednar. Pluralistic ignorance in corporate boards and firms' strategic persistence in response to low firm performance. *Administrative Science Quarterly*, 50(2):262–298, 2005.

[164] R.J. Wilson. *Introduction to Graph Theory*. Longman, 1996.

[165] Michael Wooldridge. *An Introduction to MultiAgent Systems*. Wiley, second edition, 2009.

[166] Liang Zhen and Jeremy Seligman. A logical model of the dynamics of peer pressure. *Electronic Notes in Theoretical Computer Science*, 278:275–288, 2011. Proceedings of the 7th Workshop on Methods for Modalities (M4M 2011) and the 4th Workshop on Logical Aspects of Multi-Agent Systems (LAMAS 2011).

[167] Kevin J.S. Zollman. Social network structure and the achievement of consensus. *Politics, Philosophy & Economics*, 11(1):26–44, 2014.

Samenvatting

Dit proefschrift gebruikt logische methoden om een aantal fundamentele kenmerken te onderzoeken van sociale netwerken en hun ontwikkeling door de tijd heen, waaronder het verwerken van binnenkomende informatie, en de verspreiding van opinies door een netwerk.

Deel I bevat een inleiding tot de voornaamste verschijnselen in sociale netwerken die vragen om een logische analyse van informatie en redeneren, een overzicht van achtergrondmateriaal uit de logica en sociale netwerktheorie, en een beschrijving van de hoofdlijnen van dit proefschrift.

Deel II introduceert logische modellen voor "collectief falen", en analyzeert hoe en wanneer op zich correct individueel micro-gedrag kan leiden tot contraproductief collectief macro-gedrag. Hoofdstuk 3 gebruikt dynamisch-epistemische logica's voor informatieverwerking als model voor het verschijnsel van "informatie cascades", een vorm van suboptimaal groepsgedrag. Onze analyse maakt duidelijk hoe op zich rationele actoren die anderen imiteren kunnen terechtkomen in een cascade waarbij zij de verkeerde keuze maken, ondanks de beschikbaarheid van voldoende informatie om deze fout te vermijden. We laten zien dat dit zich voordoet onder verschillende omstandigheden. Of actoren nu perfect Bayesiaans redeneren met waarschijnlijkheden of een eenvoudigere telmethode hanteren, en of zij nu onbeperkt hogere-orde kunnen redeneren over wat anderen weten en geloven of niet, sommige misleidende informatiecascades zijn domweg onvermijdelijk. Hoofdstuk 4 bestudeert een ander contraproductief sociaal verschijnsel, "pluralistic ignorance" waar iedereen zich vergist over hoe anderen denken, zoals in "de kleren van de keizer". Met behulp van hybride logica formalizeren we dit scenario en verklaren zijn dynamische eigenschappen zoals geobserveerd in de sociale wetenschappen: stabiliteit, maar ook fragiliteit. Zo blijken in alle netwerken die niet 2-kleurbaar zijn, dat een gedragsverandering van een enkele actor al volstaat om het collectieve gedrag te veranderen. Tezamen genomen bieden Hoofdstukken 3 en 4 een veelvoud aan nieuwe mechanismen voor meningsverandering van sociale actoren in gestructureerde situaties.

Deel III abstraheert van specifieke sociale scenarios naar een studie van de algemene logica achter diffusieverschijnselen in sociale netwerken, en ook van de wisselwerking tussen diffusie van meningen en de dynamiek van inkomende nieuwe informatie. Hoofdstuk 5 presenteert een algemeen raamwerk, gebaseerd op hybride logica, dat de logische wetten kan weergeven van het verloop door de tijd heen voor een grote klasse van diffusieprocessen, waarbij we een breed bereik aan mechanismen voor meningsverandering als drijvers kunnen gebruiken. Door de toevoeging van de kennislogica aan het voorgaande raamwerk onderzoekt Hoofdstuk 6 hoe de dynamiek van diffusie kan leiden tot leergedrag in actoren die waarnemen hoe publiek gedrag zich ontwikkeld in antwoord op sociale druk tot conformeren. Hoofdstuk 7, tenslotte, gaat nog een stap verder en introduceert een minimaal raamwerk om de dynamiek te modelleren van modellen waar meningen veranderen als een bepaalde drempel wordt bereikt. We laten zien hoe we nu interacties kunnen weergeven en begrijpen van topologische eigenschappen van sociale netwerken met de structuur van verschillende diffusiemechanismen. Zo tonen we bijvoorbeeld aan dat in een netwerk met voldoende dichtheid van clusters geen volledige informatie cascades kunnen optreden. Weer met een kennis-component toegevoegd laten we vervolgens zien hoe meer kennis over de structuur van een sociaal netwerk en over het gedrag van andere actoren diffusie kan versnellen in netwerken met een drempel-regel. We bestuderen ook het limietgedrag van, en bewijzen een reeks resultaten over, verschillende diffusiemechanismen met locale veranderingen die al dan niet afhangen van kennis die actoren hebben over hun situatie, waarbij we ook nog het onderscheid maken tussen kennis van de feiten en kennis over andere actoren.

Het laatste deel IV geeft een samenvatting van onze resultaten in hoofdlijnen, alsmede een vooruitblik op belangrijke verdere richtingen, gebaseerd op ons thans lopende onderzoek. We bespreken modale logica's en verwante formalismen voor de bestudering van netwerkgedrag onder verschillende topologische eigenschappen en regels voor beïnvloeding van anderen. We bespreken ook de natuurlijke overgang van verloop van netwerken door de tijd heen naar netwerkspelen waar actoren keuzes hebben en doelen nastreven.

Kort samengevat past dit proefschrift methoden toe uit moderne logica's voor informatieverandering en informatiegestuurd handelen op de studie van sociale netwerken en de verspreiding van opinies door de tijd heen, waarbij we methoden ontwikkelen die zowel gebruikt kunnen worden om specifieke sociale scenario's in detail te modelleren, maar ook om beter zicht te krijgen op de algemene redeneerwetten die ten grondslag liggen aan de dynamiek van informatie en diffusie van meningen in sociale netwerken en de processen die zich daar afspelen.

Abstract

This thesis uses logical tools to investigate a number of basic features of social networks and their evolution over time, including flow of information and spread of opinions.

Part I contains the preliminaries, including an introduction to the basic phenomena in social networks that call for a logical analysis of information and reasoning, a review of background material from logic and social network theory, plus an outline of the thesis.

Part II presents logical models of collective failures, and illuminates how and when sound individual microbehavior can lead to counterproductive collective macrobehavior. Chapter 3 uses dynamic-epistemic logics of information update to model the phenomenon of informational cascades leading to suboptimal group behavior. This analysis confirms that perfectly rational agents following the crowd may get stuck in a cascade leading them to make the wrong choice, despite the availability of enough evidence to avoid such a mistake. We show that this holds under various basic assumptions. Whether agents are full-fledged Bayesian reasoners or use a simpler counting heuristics, and whether they have unbounded higher-order reasoning or not, some misleading informational cascades are simply inescapable by rational means. Chapter 4 models a second counterproductive social phenomenon, that of pluralistic ignorance. Using a model based on hybrid logic, we formalize and explain the dynamic properties of this scenario as observed in the social sciences: its stability and its fragility. As for remedies, we show that, on all but 2-colorable network graphs, changing the behavior of one unique agent is sufficient to reverse the situation entirely. Together, Chapters 3 and 4 offer a great variety of new update mechanisms for social agents in structured settings.

Part III abstracts from specific case studies to investigate the general logic of diffusion phenomena in social networks, as well as the interaction of information and diffusion dynamics. Chapter 5 presents a general hybrid dynamic framework to capture the logical laws of the temporal evolution of a wide class of diffusion dynamics, allowing us to plug-in various network update rules. Using an epistemic

extension of this hybrid approach, Chapter 6 investigates how diffusion dynamics may induce learning by agents who observe how their public behavior evolves in response to social conformity pressure. Finally, Chapter 7 goes one step further, and proposes a minimal framework for modeling the dynamics of threshold models. We show how this setting captures interactions of network properties with diffusion processes, such as the fact that having dense enough clusters in a network prevents full cascades. Adding an epistemic logic-based component, we also show how knowing more about the network structure and the behavior of agents in the network may accelerate diffusion in threshold models. Here we study the limit behavior of various diffusion policies: knowledge-independent, first-order knowledge dependent, or higher-order knowledge dependent.

Finally, Part IV presents a summary of our findings, and some ongoing work and perspectives for future research. We discuss modal logics and related formalisms for studying network behavior under various graph properties and rules of influence. We also discuss the natural transition from network evolution by fixed rules as studied in this thesis to the study of network games where agents have choices and goals.

Overall, this thesis applies tools from current logics of information update and agency to social network analysis and opinion flow over time, offering both tools for detailed modeling of specific scenarios and a better understanding of the general laws of reasoning that underlie information and diffusion dynamics in social settings.

Titles in the ILLC Dissertation Series:

ILLC DS-2009-01: **Jakub Szymanik**
Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language

ILLC DS-2009-02: **Hartmut Fitz**
Neural Syntax

ILLC DS-2009-03: **Brian Thomas Semmes**
A Game for the Borel Functions

ILLC DS-2009-04: **Sara L. Uckelman**
Modalities in Medieval Logic

ILLC DS-2009-05: **Andreas Witzel**
Knowledge and Games: Theory and Implementation

ILLC DS-2009-06: **Chantal Bax**
Subjectivity after Wittgenstein. Wittgenstein's embodied and embedded subject and the debate about the death of man.

ILLC DS-2009-07: **Kata Balogh**
Theme with Variations. A Context-based Analysis of Focus

ILLC DS-2009-08: **Tomohiro Hoshi**
Epistemic Dynamics and Protocol Information

ILLC DS-2009-09: **Olivia Ladinig**
Temporal expectations and their violations

ILLC DS-2009-10: **Tikitu de Jager**
"Now that you mention it, I wonder…": Awareness, Attention, Assumption

ILLC DS-2009-11: **Michael Franke**
Signal to Act: Game Theory in Pragmatics

ILLC DS-2009-12: **Joel Uckelman**
More Than the Sum of Its Parts: Compact Preference Representation Over Combinatorial Domains

ILLC DS-2009-13: **Stefan Bold**
Cardinals as Ultrapowers. A Canonical Measure Analysis under the Axiom of Determinacy.

ILLC DS-2010-01: **Reut Tsarfaty**
Relational-Realizational Parsing

ILLC DS-2010-02: **Jonathan Zvesper**
Playing with Information

ILLC DS-2010-03: **Cédric Dégremont**
The Temporal Mind. Observations on the logic of belief change in interactive systems

ILLC DS-2010-04: **Daisuke Ikegami**
Games in Set Theory and Logic

ILLC DS-2010-05: **Jarmo Kontinen**
Coherence and Complexity in Fragments of Dependence Logic

ILLC DS-2010-06: **Yanjing Wang**
Epistemic Modelling and Protocol Dynamics

ILLC DS-2010-07: **Marc Staudacher**
Use theories of meaning between conventions and social norms

ILLC DS-2010-08: **Amélie Gheerbrant**
Fixed-Point Logics on Trees

ILLC DS-2010-09: **Gaëlle Fontaine**
Modal Fixpoint Logic: Some Model Theoretic Questions

ILLC DS-2010-10: **Jacob Vosmaer**
Logic, Algebra and Topology. Investigations into canonical extensions, duality theory and point-free topology.

ILLC DS-2010-11: **Nina Gierasimczuk**
Knowing One's Limits. Logical Analysis of Inductive Inference

ILLC DS-2010-12: **Martin Mose Bentzen**
Stit, Iit, and Deontic Logic for Action Types

ILLC DS-2011-01: **Wouter M. Koolen**
Combining Strategies Efficiently: High-Quality Decisions from Conflicting Advice

ILLC DS-2011-02: **Fernando Raymundo Velazquez-Quesada**
Small steps in dynamics of information

ILLC DS-2011-03: **Marijn Koolen**
The Meaning of Structure: the Value of Link Evidence for Information Retrieval

ILLC DS-2011-04: **Junte Zhang**
System Evaluation of Archival Description and Access

ILLC DS-2011-05: **Lauri Keskinen**
Characterizing All Models in Infinite Cardinalities

ILLC DS-2011-06: **Rianne Kaptein**
Effective Focused Retrieval by Exploiting Query Context and Document Structure

ILLC DS-2011-07: **Jop Briët**
Grothendieck Inequalities, Nonlocal Games and Optimization

ILLC DS-2011-08: **Stefan Minica**
Dynamic Logic of Questions

ILLC DS-2011-09: **Raul Andres Leal**
Modalities Through the Looking Glass: A study on coalgebraic modal logic and their applications

ILLC DS-2011-10: **Lena Kurzen**
Complexity in Interaction

ILLC DS-2011-11: **Gideon Borensztajn**
The neural basis of structure in language

ILLC DS-2012-01: **Federico Sangati**
Decomposing and Regenerating Syntactic Trees

ILLC DS-2012-02: **Markos Mylonakis**
Learning the Latent Structure of Translation

ILLC DS-2012-03: **Edgar José Andrade Lotero**
Models of Language: Towards a practice-based account of information in natural language

ILLC DS-2012-04: **Yurii Khomskii**
Regularity Properties and Definability in the Real Number Continuum: idealized forcing, polarized partitions, Hausdorff gaps and mad families in the projective hierarchy.

ILLC DS-2012-05: **David García Soriano**
Query-Efficient Computation in Property Testing and Learning Theory

ILLC DS-2012-06: **Dimitris Gakis**
Contextual Metaphilosophy - The Case of Wittgenstein

ILLC DS-2012-07: **Pietro Galliani**
The Dynamics of Imperfect Information

ILLC DS-2012-08: **Umberto Grandi**
Binary Aggregation with Integrity Constraints

ILLC DS-2012-09: **Wesley Halcrow Holliday**
Knowing What Follows: Epistemic Closure and Epistemic Logic

ILLC DS-2012-10: **Jeremy Meyers**
Locations, Bodies, and Sets: A model theoretic investigation into nominalistic mereologies

ILLC DS-2012-11: **Floor Sietsma**
Logics of Communication and Knowledge

ILLC DS-2012-12: **Joris Dormans**
Engineering emergence: applied theory for game design

ILLC DS-2013-01: **Simon Pauw**
Size Matters: Grounding Quantifiers in Spatial Perception

ILLC DS-2013-02: **Virginie Fiutek**
Playing with Knowledge and Belief

ILLC DS-2013-03: **Giannicola Scarpa**
Quantum entanglement in non-local games, graph parameters and zero-error information theory

ILLC DS-2014-01: **Machiel Keestra**
Sculpting the Space of Actions. Explaining Human Action by Integrating Intentions and Mechanisms

ILLC DS-2014-02: **Thomas Icard**
The Algorithmic Mind: A Study of Inference in Action

ILLC DS-2014-03: **Harald A. Bastiaanse**
Very, Many, Small, Penguins

ILLC DS-2014-04: **Ben Rodenhäuser**
A Matter of Trust: Dynamic Attitudes in Epistemic Logic

ILLC DS-2015-01: **María Inés Crespo**
Affecting Meaning. Subjectivity and evaluativity in gradable adjectives.

ILLC DS-2015-02: **Mathias Winther Madsen**
The Kid, the Clerk, and the Gambler - Critical Studies in Statistics and Cognitive Science

ILLC DS-2015-03: **Shengyang Zhong**
Orthogonality and Quantum Geometry: Towards a Relational Reconstruction of Quantum Theory

ILLC DS-2015-04: **Sumit Sourabh**
Correspondence and Canonicity in Non-Classical Logic

ILLC DS-2015-05: **Facundo Carreiro**
Fragments of Fixpoint Logics: Automata and Expressiveness

ILLC DS-2016-01: **Ivano Ciardelli**
Questions in Logic