A Unifying Semantics for Causal Reasoning about Action

by

Mikhail Prokopenko

A thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy

Computing Department
Division of Information and Communication Sciences
Macquarie University
Sydney, Australia

August 2001

Statement

Apart from standard techniques and results, the use of the work of other authors has been indicated by references in the text. Parts of chapters 3, 4, 5 and 6 have been published under joint authorship with my supervisors Doctor Maurice Pagnucco, Doctor Pavlos Peppas and associate supervisor Doctor Abhaya C. Nayak as indicated. In conformity with the regulations of the Macquarie University it is claimed that the work presented in this dissertation has not been submitted for a higher degree to any other university or institution.

Acknowledgements

This dissertation would not have been possible without the kind assistance and support of many people, who have contributed to this work through technical or philosophical discussions of different parts of it, or simply through their encouragement. The influence of my Ph.D. advisors on the presented work is second to none. Pavlos Peppas introduced me to the research fields of Reasoning about Action and Belief Change, and outlined challenging and inspirational goals and strategies. Maurice Pagnucco and Abhaya C. Nayak provided very generous and unswerving guidance on the journey, charting the way through various non-monotonic logics, syntactical and semantical traps, peculiarities of causal knowledge and deceptive ramifications of seemingly rational choices. I have never ceased to be fascinated by enlightening clarity, elegant precision and philosophic depth of the research carried out by my Ph.D. advisors. It is a pleasure to acknowledge their supervision and to thank them for it.

I would like to thank a number of other people who have provided useful input on topics related to the dissertation. I have had the good fortune to be associated with Norman Foo, the leader of the Knowledge Systems Group at the University of New South Wales, whose insights into logic-based approaches to systems modelling have been invaluable. I also had the privilege of discussing a few important problems in causal reasoning with Eric Sandewall. It would be impractical to list all of my friends and colleagues who have contributed to my understanding of non-monotonic and temporal reasoning through collaborative research efforts or informal albeit very fruitful discussions. I may only single out a few from among the many — my thanks to Aditya K. Ghose, Yan Zhang, Chitta Baral, Michael Gelfond and Abdul Sattar. I would also like to thank the anonymous referees for valuable suggestions.

I am indebted to the Commonwealth Scientific and Industrial Research Organisation (CSIRO) for years of support. At CSIRO, I am especially grateful to Rhys Francis, Ryszard Kowalczyk, Craig Lindley and Cécile Paris for their encouragement, and for giving me the opportunity and time to complete this thesis. Special thanks to my colleagues Marc Butler, Thomas Howard, Victor Jauregui and Peter Wang — their uncompromising pursuit of knowledge and perfection not only has given me confidence and cheered me up on various occasions over the last years, but continues to inspire me.

Let me also express my sincere gratitude to the teachers and advisors who have influenced me so profoundly throughout the years of my education. I am especially thankful to Naum G. Chernoguz, Alek S. Samedov and Ramiz J. Aliyev who played a vital role in shaping my interests in artificial intelligence, mathematical logic and functional analysis.

I thank my Australian and overseas friends who continued to encourage me regardless of distances and events separating us. I wish to offer a special word of gratitude to my mother Berta Vysotskaia and grandmother Elizaveta Kahn for their moral support and the selfless sacrifices that they made in allowing me to pursue my education. Finally, I thank my wife Elena and our daughter Oksana for their love and patience, and for being with me during all the turbulent years of our odyssey. Here you are...

Abstract

One of the principal concerns in the research area of Reasoning about Action and Change is determining the ramifications of actions in changing environments. A particular tendency emerging in recent literature endorses the explicit incorporation of causal knowledge in logic-based action theories. It is argued that causal extensions not only enhance the expressive power of theories of action, but may also provide more concise and intuitive representations.

This dissertation investigates semantics for causal reasoning about action and change. It does so by exploring the role of several fundamental underlying principles, such as the *Principle of Minimal Change* and the *Principle of Causal Change*. The development of this work culminates in a general unifying semantics for a class of action theories represented by a number of recent and influential approaches.

We focus on three of the most prominent causal frameworks in the Reasoning about Actions literature: the causal systems with fixed-points suggested by McCain and Turner, the causal relationship approach of Thielscher, and Sandewall's causal propagation semantics (also known as the transition cascade semantics). Each is studied via a semantics which includes a preferential component augmented with a causal relation.

The foregoing results are used to develop *a general augmented preferential-style semantics* that subsumes the causal systems with fixed-points and the causal relationship approach. The causal propagation semantics of Sandewall is shown to be a special case as well, characterised under certain uniformity assumptions.

The unifying general augmented preferential semantics, emerging as a result of this study, captures both Principles of Change and shows their clear and distinct roles — they are not inter-reducible but go hand-in-hand in causal action theories. Furthermore, the general semantics emphasises the role of contextual information affecting both minimality and causality, and provides a means for balancing different contributing factors. It is argued that hidden or less immediate forces shaping our motivating approaches become transparent with the help of the general semantics. In addition, it is hoped that the unifying semantics would provide further insights into views on causation and minimality, shared by these and other approaches.

Contents

Chapter 1

Introduction

Reasoning about Action and Change is one of the most intriguing and fundamental issues in Artificial Intelligence. An intelligent agent, be it a robot, cyborg or synthetic software agent (softbot), is expected to interact with its environment and reason about the interactions. Sometimes, the effects of an agent's actions can be traced relatively easily. On other occasions, an action may result in intricate and convoluted ramifications. Arguably, agents' ability to reason about direct and indirect effects of actions is a distinguishing feature of intelligence. Ultimately, agents' existence and survival in the environment depends on their competence in reasoning about changes in the environment.

Reasoning about actions and change may take many forms. For example, behaviour of simple biological organisms and basic situated artificial agents embeds reasoning about change in low level reactions. In particular, tropistic (and hysteretic) reactions map sensory inputs (and the agent's internal state) into particular actions available to the agent, producing an adequate reactive behaviour [16, 49]. More complex life forms (natural or synthetic) are able to represent the environment, model it and reason about consequences of their actions. One fundamental characteristic of such representations is an *explicit* notion of change, or in other words, an incorporation of "time's arrow" (the temporal asymmetry[1]).

Let us briefly illustrate this point with reasoning about basic movement. While an

[1]For example, an agent may consider that events depend on earlier events in a way in which they do not depend on later events, and subjectively deliberate for the future on the basis of information about the past.

ability to detect a change in direction may be deeply embedded at a sensory level (there are direction-sensitive retina cells in rabbits' eyes, for example), a measurement of a shift in observed positions would seem to require a rudimentary reasoning process, dealing with at least two time points, states of affairs or events. Furthermore, a notion of object velocity emerges after a series of measurements, and becomes a powerful tool in modelling the dynamics of the world.

Basically, an agent becomes *situated in time* in addition to being *situated in space* before it develops a higher reasoning level. Not unlike the counting ability that grows into arithmetic and then algebra, some convention on "time's arrow" underlies increasing levels of reasoning about change.

It is well recognised that intelligent tasks such as prediction, planning, explanation assume some distinctions between the past, the present and the future and involve some form of temporal reasoning. Obviously, temporal reasoning is not restricted to reasoning about time itself. It "also includes reasoning about phenomena that take place in time, i.e. *reasoning about actions and change*" [55]. Whether an artificial agent is expected to calculate a moving object's position over time, determine the state of an electric circuit, or find out a reason for the fire that destroyed a house, it must assume (among others) some notion of change, action and causation. Ideally, if reasoning is expected to be consistent and systematic, these notions should be formalised, leading to reproducible and comparable results across agents. In other words, an intelligent agent has to incorporate a formal reasoning system that produces inferences about the effects of actions.

The Reasoning about Action research area primarily investigates formal logic-based approaches describing the effects of an agent's actions on the environment. It is important to realise, however, that a unique and completely general-purpose logic of reasoning about action and change is no longer perceived as the main objective. It has been argued that "perhaps the logic of common-sense reasoning, rather than being unified and concise, will have the character of a Swiss army knife and contain one tool for each purpose" [55]. In other words, various reasoning systems may be based on different theories of action. This highlights the role of a systematic methodology that allows us to analyse and compare action theories — with respect to some underlying semantics.

One particular trait emerging in recent literature on Reasoning about Action attempts

to explicitly embody a notion of causality (causation) in logic-based action theories[2]. It is argued that such an extension would not only enhance the expressive power of theories of action, but may also provide more concise representations [37, 33, 63]. What seems to be lacking so far is a general semantic framework that covers this particular class of action theories.

This dissertation is a semantic investigation into the role of causality in reasoning about action and change. It attempts to examine some of the aspects of "time's arrow" and causality, explores the role of several important underlying principles, and introduces a general semantics for a class of action theories represented by a number of recent and influential approaches.

1.1 Preliminaries and Background

The area of Reasoning about Action has grown considerably in the last decades, and overlaps now with many other fields, as diverse as philosophy of time and causation and robotic soccer. This can be partially explained by the fact that many related, though distinct, areas share some essential problems, crystallised and investigated under the heading of Reasoning about Action (such as planning, explanation, prediction).

Following Sandewall and Shoham [55] we say that a reasoning task typically involves "(1) designation of certain actions which have been (will be, may be) performed, as well as their order of execution; (2) statements about the state of the world before the actions; (3) statements about the state of the world after the actions." In a planning task, (2) and (3) are given and (1) is sought, while a prediction task uses (1) and (2) in determining (3).

A more general interpretation is the (extended) prediction problem, referred to by Shoham [59] as a problem of "how to reason *efficiently* about what is true over extended periods of time", while maintaining "certain tradeoffs between risk avoidance and economy":

> The most conservative prediction refers to a very short interval of time,

[2]Here, by "causality" we mean a $category$ of causal connection (causation), rather than a $principle$ (the general law of causation) stating the form of the causation, or a $doctrine$ of causal determinism (causalism) asserting that everything happens according to the causal law [5, p. 3].

> in fact an instantaneous one, but that makes it very hard to reason about more lengthy future periods. For example, if on the basis of observing a ball rolling we predict that it will roll just a little bit further, in order to predict that it will roll a long distance we must iterate this process many times (in fact, an infinite number of times).
>
> The disadvantages of the conservative prediction which refers to only a short time period suggest making predictions about more lengthy intervals. For example, when . . . you throw a ball into the air you predict that it will have a parabolic trajectory. The problem with these more ambitious predictions is again that they are defeasible, since, for example, a neighbor's window might prevent the ball from completing the parabola."

In summary, the general extended prediction problem is that an agent needs to make a lot of predictions about short future intervals before predicting something about the more distant future.

It is interesting to note that two challenging problems in Reasoning about Action — the *Frame* and *Ramification* problems — are related to (and arguably, can be subsumed by) the extended Prediction problem. Informally, the Frame problem is concerned with what does not change when an event occurs or an action is performed. Sometimes, the term "Frame problem" is given a broader scope, but typically it is used in the restricted sense of the *Persistence* problem: assuming that properties of the world do not change unless affected by an action (an event), the aim is to build a reasoning system that models the dynamics of the world in an efficient and convenient (concise) way.

The Persistence problem may seem to be quite artificial or at least formalism-specific; however, difficulties arise when an agent is faced with complex indirect effects of actions. In this case, it is not sufficient just to update directly affected properties and leave the rest unchanged — some action consequences may spread quite far and affect seemingly remote and unrelated properties (for instance, the "domino effect" scenario). In other words, the agent also faces the *Ramification* problem — how to formalise all of the things that do change as the result of an action. Ginsberg and Smith [17] describe the problem as follows:

> The difficulty is that it is unreasonable to explicitly record all of the consequences of an action, even the immediate ones. ... For any given action, there are essentially an infinite number of possible consequences that depend upon the details of the situation in which the action occurs.

It is precisely a combination of the Frame and Ramification problems that makes a search for a concise solution extremely challenging and non-trivial.

Typically, the Reasoning about Action tradition suggests to represent domain knowledge declaratively in a formal language capable of inferring "that a certain strategy will achieve its assigned goal" [39]. Logic has traditionally been chosen as the representation language and various reasoning systems have been designed to address the Frame and/or Ramification problems: situation calculus [39], default logic [52], circumscription [40, 27], temporal logic of chronological ignorance [59, 60], action languages [18, 23, 19, 31, 67], fluent calculus [63], features and fluents framework [54] procedures such as STRIPS [11], the Possible Worlds Approach (PWA) [17], the Possible Models Approach (PMA) [70], causal fixed-points [37], causal relationships approach [62, 63], etc. Historically, there is an agreement that "the 'knowledge content' of a reasoning program ought to be represented by data structures interpretable as logical formulas of some kind" [53].

We shall postpone a formal description of our motivating approaches till sections 2.7 — 2.9, and shall try to use, in this section, only informal allusions to various concepts of action theories. This will allow us to highlight these concepts and the underlying principles in a natural and intuitive way, thus clarifying our main objective — a general semantics for a class of action theories embodying causality.

Usually, most solutions require that action specifications provide *direct* (most significant, immediate, etc.) *effects* explicitly, and employ domain constraints of some form for specifying additional (indirect) changes that may occur due to the action.

However, there are different methodologies for determining which propositions hold after performing an action. The monotonic situation calculus and some non-monotonic logics (default logic, circumscription, logic of chronological ignorance, action languages) try to infer what propositions are true once the events have occurred (actions have been

performed), and thus answer queries about the theory without actually updating it. Following McCarthy and Hayes [39]: the facts about situations are "used to deduce further facts about that situation, about future situations and about situations that persons can bring about from that situation". An alternative way to formalise reasoning about change was proposed in the STRIPS approach [11] and extended in the PWA and the PMA. "The basic insight ... is that the world does not change much from one instant to the next" [17]. So it is possible to maintain a current state of the world and incorporate an update procedure constructing "the nearest world to the current one in which the consequences of the actions under consideration hold".

In other words, an agent follows the *Principle of Minimal Change*. According to this principle, the world changes as little as *possible* when an action is performed. Basically, this principle enables the agent to reduce the amount of explicit information about what changes and what persists through an action. While the metaphysical status of the Principle of Minimal Change is an unsettled topic (that will be briefly discussed later), a precise definition of minimal change depends on the particular formalism in question. Often, it is defined by set inclusion, and presumes that the total set of changes resulting from an action contains those changes that are explicitly specified as direct effects of the action, and a *minimal* set of other changes required by the domain constraints. Sometimes, a particular measure of minimal change assigns different degrees of inertia to properties under consideration (a policy of *categorisation*), which allows a reasoning system to assume persistence for more basic (independent) properties and apply domain constraints to secondary (derived) properties. In general, an agent uses a preference relation in accepting outcomes (states, sets of states, interpretations, models) that are strictly closer to the initial one than other possibilities (which are rejected).

In addition, some action theories embody background information in the form of domain "causal rules" or constraints, and apply the *Principle of Causal Change*. Informally, these approaches specify how changes in one property (state variable, state of affairs, event, state) may "cause" changes in another. Then, a reasoning system is expected to produce (in response to an action) the outcome which satisfies the action's direct effects and the domain constraints, while incorporating only justified changes. Intuitively speaking, all the accepted properties (states of affairs) must be justified by the

underlying causal constraints.

Sometimes, the Principles of Minimal and Causal Change are applied together, resulting in policies of causal minimisation. Let us illustrate a variant of such a policy with an informal example borrowed from Umberto Eco's "The Name of the Rose" [9, p. 91] — the example is concerned with possible causes of Adelmo the librarian's death[3]. The puzzle is that Adelmo's corpse, lacerated by rocks, is found in a heap of straw below a high (east) tower.

> "And so, think whether it is not less — how shall I say it? — less costly for our minds to believe that Adelmo, for reasons yet to be ascertained, threw himself of his own will from the parapet of the wall, struck the rocks, and, dead, wounded as he may have been, sank into the straw. Then the landslide, caused by the storm that night, carried the straw and part of the terrain and the poor young man's body down below the east tower."
>
> "Why do you say this solution is less costly for our minds?"
>
> "Dear Adso, one should not multiply explanations and causes unless it is strictly necessary. If Adelmo fell from the east tower, he must have got into the library, someone must have first struck him so he would offer no resistance, and then this person must have found a way of climbing up to the window with a lifeless body on his back, opening it, and pitching the hapless monk down. But with my hypothesis we need only Adelmo, his decision, and a shift of some land. Everything is explained, using a smaller number of causes."

In other words, an agent reasons that the world changes as little as *necessary* when an action is performed. One particular approach following this kind of causal minimisation is McCain and Turner's approach [37] that introduces *causal fixed-points*. Intuitively, a causal fixed-point is an outcome incorporating the direct effects of actions, where all other changed properties are causally justified (in a certain sense). In other words, every

[3] We follow here a rich tradition of assassination examples in the area of Reasoning about Action, and believe that this one may serve as a good illustration of a (not necessarily precise) causal minimisation policy.

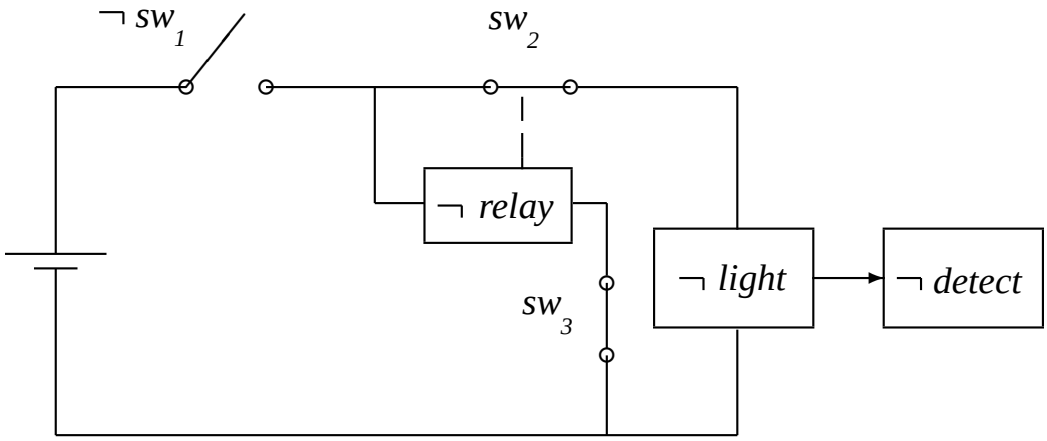

Figure 1.1: The electric circuit with Light Detector.

detail in the outcome must be "explained" either as persisting through the action, or as a direct effect, or as a causal ramification of *other properties contained in the outcome* — hence the fixed-point flavour. Obviously, a possible outcome that contains at least one detail without such justifications is rejected, even if it does not violate the domain constraints. It is not hard to observe that causal fixed-points, indeed, incorporate only necessary changes.

It has been argued in some recent proposals that the Principle of Minimal Change can be replaced or subsumed by the Principle of Causal Change in reasoning about actions: "the aim of generating ramifications is not to minimize change but to avoid changes that are not caused, which ... need not be identical" [63]. These approaches allow a reasoning system to propagate beyond just nearest possible states towards states where all changes are justified. One interesting example is the Light Detector example, proposed by Thielscher [63]. This example illustrates that sometimes one possible successor to an initial state of the world may have strictly more changes than another successor, while both of them seem to be intuitive.

Delaying a formal description of the example for subsequent sections, we merely sketch it here. An electric circuit includes three switches, a relay, a light bulb and a light detector device (Figure 1.1). Initially, both the light bulb and the detector are off. The circuit is specified in such a way that it is possible (by toggling one of the switches)

to activate both a relay and a sub-circuit involving the light bulb for an instant — before another switch jumps its position as a ramification of activating the relay, and turns the light off. It is argued in [63] that at this brief instant the detector may react to the light. Despite the fact that the light bulb itself does not stay activated, the detector may. Therefore, two outcomes are presented as possible: one where the light is off, and the detector is not activated, and another, where the light is off as well, but the detector is activated. Obviously, the second outcome has strictly more changes (with respect to the initial state) than the first — the change in the detector's state is an extra ramification. Importantly, this change is justified "during" some dynamic process employing causality (eg., propagating from "no light and no detector" to "light and no detector" to "light and detector" to "no light and detector"). Unlike fixed-point details, the presence of the activated detector is justified not by other "contemporary" properties (statically) contained in the outcome, but rather by some (dynamic) propagation of change.

The *causal relationship approach* described in [63] formalises a proposal capturing both successors in this example, and argues that the Principle of Minimal Change "is not always adequate for distinguishing between possible indirect effects on one hand, and unfounded changes on the other hand" [63].

As an aside, it should be noted that this type of "causal" reasoning may still be implicitly based on some preference relation. For example, an agent may "prefer sequences of world states in which one world state leads causally to the next, rather than sequences in which one world state follows another at random and without causal connection" [61].

We do not intend here to discuss the questions on how intuitive the "non-minimal" successor is in the Light Detector example, and whether there are some hidden (temporal) dependencies in the specified circuit. What seems to be more important, however, is a relationship between the Principle of Minimal Change and the Principle of Causal Change. Can the former indeed be subsumed by the latter? Are they inter-reducible?

Not surprisingly, the ontological status of the Principle of Causal Change is unclear at least as much as that of Minimal Change. For a long time it has been grounded in an open-ended debate on the metaphysics of Causation. This subject will be revisited in the following chapters. At this stage we simply argue that it appears to be extremely hard to compare the roles of the two principles of change *within* particular action theories. Not

only are there different interpretations across the field, but the principles' manifestations are often limited by the operational mechanics of particular reasoning approaches. One attractive option is to offer a general semantics for a class of action theories and explore the roles of the underlying principles, highlighting different perspectives on "time's arrow".

1.2 Overview

While this work is not intended to shed new light on the metaphysics of causation and minimality, it aims to investigate the common ground taken by different approaches to reasoning about action and change. In Chapter 2, we shall attempt to set a framework for a general semantics, relying on certain fundamental principles (such as the Principle of Minimal Change and the Principle of Causal Change), and clarify the reasons that allow us to hope that our motivating approaches can be represented in a unifying setting.

As a next step, we intend to study possible areas of interaction between diverse characteristics of action domains, such as inertia and causality, in the context of action languages with different commitments towards causality and categorisation policies [18, 23, 31, 67]. This investigation, in Chapter 3, is not related directly to our quest towards a unifying semantics, but is a necessary step permitting us to dismiss a particular categorisation policy in action theories operating with causality.

In the two following chapters, 4 and 5, we shall attempt to provide a preferential-style semantics (augmented with a causal relation on states) for our motivating approaches: the causal fixed-points framework of McCain and Turner [37] and the causal relationship approach of Thielscher [62, 63]. The attempts will reduce gaps between the approaches and draw attention to remaining (contextual) differences.

Then, in Chapter 6, a variant of the augmented preferential semantics will be related to the causal propagation semantics (the transition cascade semantics) of Sandewall [56, 57], subsuming it under certain assumptions. Here, the primary target would be to discover the role of minimality in action invocations.

Finally, Chapter 7 will introduce a general augmented preferential semantics, and summarise the reductions obtained in the earlier chapters. The general semantics will

emphasise the role of contextual information affecting both minimality and causality, and provide a means for balancing competing and collaborating factors. Discovering hidden or less immediate forces shaping our motivating approaches, would allow us to make their differences transparent in the general semantics.

The unifying semantics, emerging as a result of this study, captures both principles of change and shows their clear and distinct roles. They are not inter-reducible but go hand-in-hand in causal action theories. This may indicate that both principles are required to solve the Frame and Ramification problems in a *concise* fashion.

Chapter 2

Framework

An agent reasoning about action and change may represent a dynamic world in many ways, choosing certain components and discarding others. In this chapter we shall discuss different aspects of world dynamics and its representations, while trying to develop our framework incrementally.

2.1 Fluents, States and Actions

How can an agent represent change? What are the aspects of temporal and causal asymmetries[1] that the agent may perceive, represent and reason about? These questions lie at the very core of Reasoning about Action, and not surprisingly, are typically answered from very different philosophical viewpoints. Essentially, "almost *any* change can be thought of both historically (in terms of sequence of states, i.e., a change of state) and experimentally (as a new *kind* of state, a state of change)" [12].

Let us begin with the first viewpoint. Representing actions and change in terms of states and state-transitions is a well-established tradition in Artificial Intelligence. Informally, "a state is a snapshot of the underlying dynamic system, i.e. the part of the world being modeled, at a particular instant of time" [63].

Sometimes, a state is completely described by all relevant facts about it — the intensional view, that refers to "internal structure of world states in terms of objects and

[1] As mentioned in [47], "There are a number of apparently distinct ways in which the world we inhabit seems asymmetric in time. One of the tasks of an account of temporal asymmetry is thus a kind of taxonomic one: that of cataloging the different asymmetries (or 'arrows', as they have come to be called), and sorting out their family relationships."

relations specific to the representation scheme at hand" [44]. In this case, state variables or *features* become basic primitives. Features may be formed constructively by starting with a finite number of individually named objects [58], and used in conjunction with *fluents* that represent functions from features to values (for example, truth-values, real numbers or integers). Occasionally, the term "fluent" is used to mean "feature", similarly to a function being identified with its symbol.

Some authors treat states of the whole universe as uninterpreted points, and without necessarily describing their internal details — the extensional view, that "does not rely on the choice of the specification language, which can only be achieved if we model properties without referring to the internal structure of world states" [44].

In either case, a "world-state" can be assigned to a time-point. Then it could be argued that "change arises as a by-product of the assignment of states to times" [12], meaning that the history of the world is a (temporally) ordered set of states.

Our goal is to describe a general semantics for a class of action theories. We are bound, therefore, to take the extensional view and avoid assumptions about the internal structure of world states. More precisely, we shall denote the set of all world states defined for a specific representation scheme[2] by $\mathcal{W}$, and consider a world state as an uninterpreted point in the space $\mathcal{W}$. While it may sound as a limitation on a class of potentially definable action theories, we intend to demonstrate that most of the crucial concepts can be captured in this representation-independent style.

The state-based account of change is not entirely unavoidable. An alternative would be to postulate a notion of event which is distinct from and independent of the notion of a state. This can be taken to the extreme by recognising only those states which can be characterised in terms of events — an event-based account of change. More realistic, however, is "a *mixed* account, in which both states and events are admitted as primitive terms. In such a model, one has to specify the logical and causal relationships which hold between states and events" [12].

Another interesting refinement is a distinction between natural events and volitional actions (that involve free-will based decisions). Both events and actions may lead to changes of world states, and it was argued that sometimes the distinction is necessary

[2]For our purposes, it is sufficient to consider this set to be a finite set.

to improve event minimisation strategies aiming to differentiate between caused and unmotivated event occurrences [64]. However, we would make too strong an assumption by splitting events into these two types (volitional and natural) before exploring a weaker option. More precisely, following the mixed account of change, we introduce a finite set of events (or actions) $\mathcal{E}$, without speculating about the internal structure or type of the events.

While precise specifications of logical and causal relationships which hold between states and events belong to a particular representation scheme, we now need to indicate how event (action) effects can be reflected in the state-space $\mathcal{W}$. Arguably, without such a link it is not possible to express much about state changes. One natural way is to introduce the post-condition of an action $e \in \mathcal{E}$ as the property that e directly brings about with its occurrence. We shall denote the post-conditions of the action e by $[e]$. Since we do not wish to commit ourselves to particular representation details, we represent $[e]$ as a subset of $\mathcal{W}$. Intuitively, the post-condition of e is precisely the properties common to all states in $[e]$. It is important to realise that post-conditions $[e]$ are not made conditional on the initial states where the action may be executed, and therefore, are captured unvaryingly and uniformly by a subset of $\mathcal{W}$. In other words, whenever the action e is performed, an agent considers states in the set $[e]$ as states compatible with the action's direct effects. Formally, we define $[e]$ to be a function from $\mathcal{E}$ to $2^{\mathcal{W}}$ (the power-set of $\mathcal{W}$), such that for all actions e in $\mathcal{E}$,

$$[e] \subseteq \mathcal{W}, \;\; [e] \neq \emptyset.$$

Our choice to define the effects of actions via states follows an entirely extensional approach, while an intensional alternative would be to specify which fluents (state variables) must change in order to incorporate action post-conditions. This is not surprising given our objective — a general semantics for a class of action theories.

In summary, in introducing $\mathcal{W}$, $\mathcal{E}$, and $[e]$ we follow a mixed account of change (state-based and event-based), while staying totally within the extensional approach. The only assumptions we have made so far are that state-space and action-space are non-empty, and any action's post-condition is a non-empty subset of the state-space (i.e., it is satisfied by at least one state).

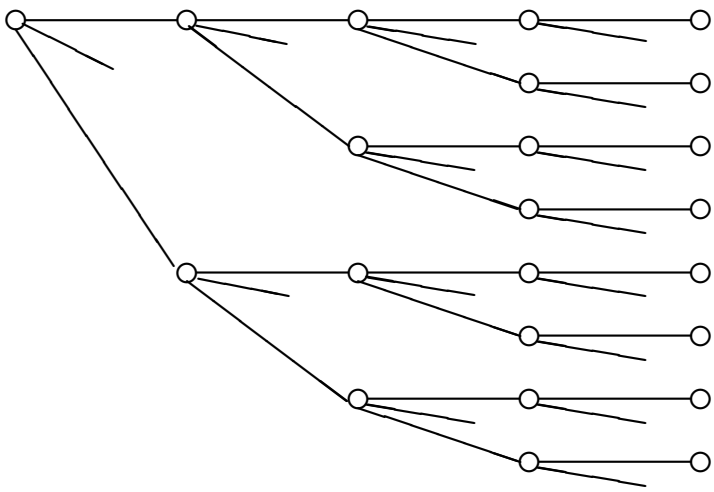

Figure 2.1: Branching tree of state transitions.

2.2 Possible, Legitimate and Successor States

2.2.1 Domain Constraints and State Transitions

The state-space $\mathcal{W}$ contains, in principle, all *conceivable* states of an underlying dynamic world (system). Given a particular representation scheme and a choice of state variables (fluents or state functions), one may form a state-space intensionally (in other words, by referring to the internal structure of states) — for example, as "the Cartesian product of the finite range sets of a finite number of state variables" [56]. In other words, each conceivable combination of variable components is treated as a possible state, and together they make up the conceivable state-space.

This view makes use of the idea of a model of the world that satisfies the requirements of *Logical Atomism* [69, p. 110]. There is a set of n basic features (or states of affairs), and a state of the world, at any given time, is a conjunction with n terms such that each of the basic features or its negation appears as a term. Hence, there are 2^n states that are logically possible. Given a sequence of k events (actions) or "occasions" following von Wright's terminology [69, p. 108], forcing state transitions, the number of all possible successions (histories) of the world is 2^{kn}. The possibilities of state transitions can be depicted in a topological figure, as a branching tree (Figure 2.1).

However, since the basic components may be inter-related and mutually restricted, not every combination represents a nomologically (lawfully) possible state. This conjecture definitely presupposes the existence of underlying laws [6]:

> Only those values of the components of the total state function that are compatible with the laws will be really (not just conceptually) possible. In other words, because the laws impose restrictions upon the state functions and their values, hence upon the state spaces, only certain subsets of the latter are accessible to the thing represented. We shall call the accessible part of the state-space the *lawful state space* of the thing in the given representation and relative to a given frame.

In short, the lawful state-space is a *proper* subset of the state-space $\mathcal{W}$. The elements of this subset (lawful states) are sometimes referred to as *admitted* [56] or *legitimate* [44] states. In the context of many logics of action, legitimate world states are defined as the elements of $\mathcal{W}$ satisfying certain conditions known as *domain constraints*. The domain constraints are often specified through given fluents and syntax-dependent relations, and therefore, shall not be used directly in our representation-independent, extensional approach.

Instead, we introduce the set $\mathcal{D}$ of legitimate states explicitly — as a proper nonempty subset of $\mathcal{W}$ (Figure 2.2). Ideally, given a particular scheme representing domain constraints, our semantics will identify the elements of the set $\mathcal{D}$ with the states satisfying these constraints. Importantly, the semantics is not reliant on what kind of constraints (logical, functional, causal, etc.) makes a particular state illegitimate (unlawful or inadmissible). In other words, the forces that shape the internal structure of a world state may require specific representations, obscuring our search for weakest semantical assumptions, covering more classes of action domains.

We have committed so far to saying that not every element of $\mathcal{W}$ represents a nomologically possible state. Moreover, not every state transition is a possible development. Again, there may be multiple reasons for this, dependent on underlying laws. A discussion on the nature of such laws (functional, causal and so on) is outside the scope of this work, and we will only briefly address this issue in Section 2.5.3. At this stage, we just

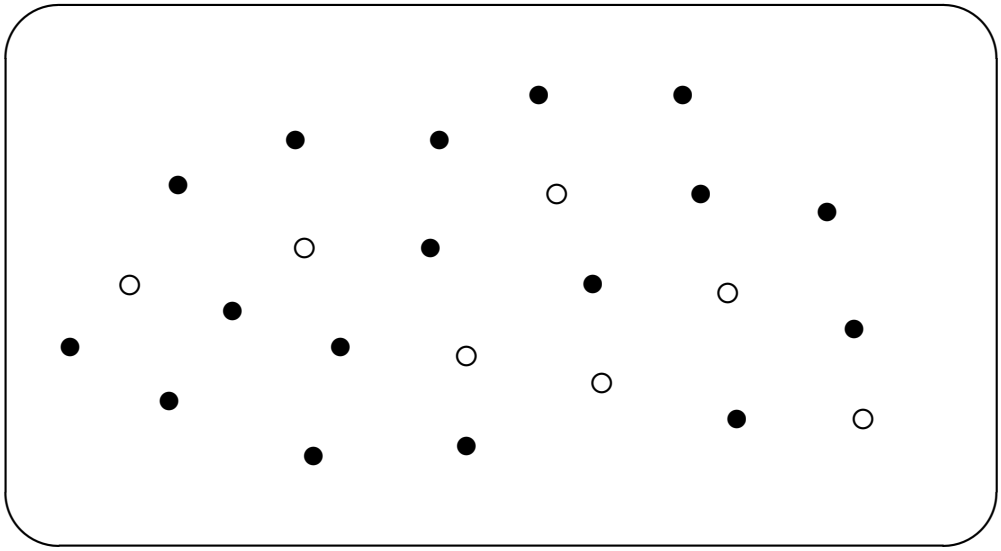

Figure 2.2: The legitimate states (members of the set $\mathcal{D}$) are depicted as filled circles.

point out that the number of all really possible state transitions for k actions is less than the number 2^{kn} of all conceivable histories.

What is needed, therefore, is a concise way to specify the states resulting from an action $e \in \mathcal{E}$ executed at an initial state $w \in \mathcal{W}$ (or more precisely, at a state $w \in \mathcal{D}$, as we do not wish to entertain the possibility of being in an illegitimate state in the first place — actually, we are not concerned with what happens at illegitimate states). In other words, we need to describe the *successor* state(s), where the underlying system moves to as a result of an action e performed at a state w. Formally, we define a *selection* or *result* function $Res(w, e)$ to be a function from $\mathcal{W} \times \mathcal{E}$ to $2^{\mathcal{W}}$, mapping a state w and an action e to a set of (legitimate) successor states.

An example of the selection function is a simple function choosing all legitimate states compatible with the action's direct effects:

$$Res_*(w, e) = \mathcal{D} \cap [e].$$

Obviously, this particular function would not, normally, satisfy an intelligent agent engaged in reasoning about actions — as successor states may be chosen without any connection to the initial state w, shown as the left-most circle in Figure 2.3.

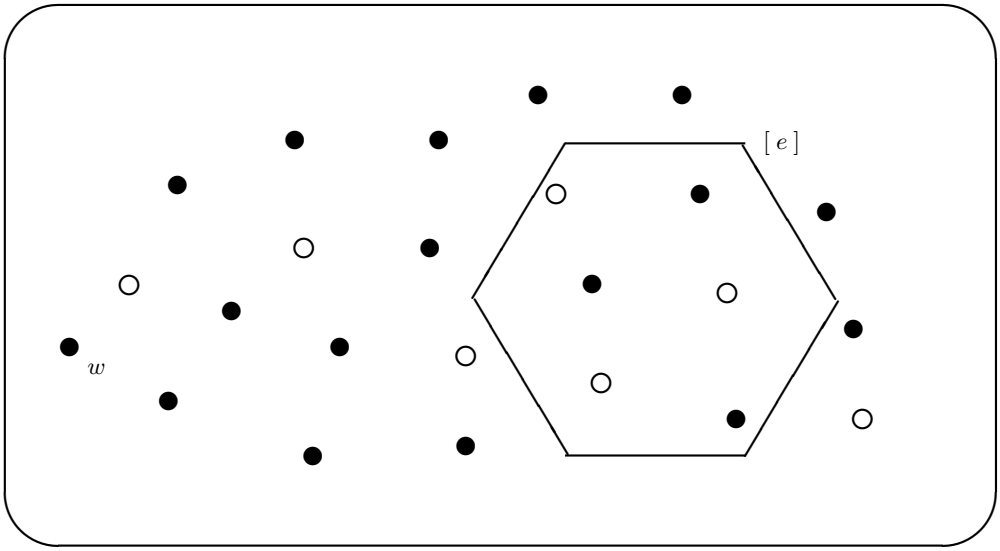

Figure 2.3: The states compatible with direct effects of the action e are enclosed within the polygon area $[e]$, and the simple selection function $Res_*(w, e)$ chooses all three states depicted as filled circles, inside the polygon.

2.2.2 Systematic Methodology

Ideally, the definition of the function $Res(w, e)$ must be intuitively convincing since it is intended to capture common-sense notions of system dynamics. Not only should it be plausible semantically (in an abstract sense), but also it must allow us to verify it with concrete examples. The latter objective has been quite vigorously pursued in Reasoning about Action research. However, the hope that a number of indicative examples of common-sense inference will facilitate an extraction of a common-sense reasoning logic, applicable to many other "similar" examples, has faded with time. In short,

> ...the methodology of common-sense examples has resulted in a somewhat chaotic development: logics, examples, and counter-examples have been confronted, and it has not always been clear which property of logic was to be given the credit for its successes, or the blame for its failures. It has not even always been clear what was a success or a failure in terms of the proposed reasoning examples [55].

In other words, the selected examples may not reveal all the aspects of the reasoning task under consideration. Moreover,

> ...even if one has very strong intuition and chooses a set of representative examples that covers every aspect of the problem, leading to the development of a logic that works for the general case, we will not have any means of *formally proving the correctness* of the proposed logic; with the methodology of representative examples one can only be proved wrong [44].

The *systematic methodology* proposed by Sandewall [54] suggests, instead, to progressively investigate classes of domains, providing each one with an applicable action theory. For each class we may build an abstraction (a semantics) specifying, for example, how to obtain successor states or which logical models are intended [55]. Then, we may design a reasoning system that operates according to the specification, in a provably correct way.

Let us briefly illustrate the systematic approach in the context of first-order logic. Let Δ be a set of formulae, and let $\Phi(\Delta)$ denote the set of classical models for Δ. The set of intended models, denoted as $\Sigma(\Delta)$, is then specified as a subset of $\Phi(\Delta)$. This leads to the question of how to obtain $\Sigma(\Delta)$ or the corresponding *conclusions* (formulae that are true in all members of $\Sigma(\Delta)$) in terms of operations on formulae in Δ [55]. Once Σ is defined, one may introduce a variety of entailment methods defined as functions from a set of axioms Δ to a set of models $\Upsilon(\Delta)$, and pose questions as to correctness, soundness and completenes of a particular entailment method $\Upsilon(\Delta)$ with respect to $\Sigma(\Delta)$. More precisely, following [55]:

- For a given entailment method Υ and a given Σ, does the method obtain exactly the intended models, i.e., is $\Upsilon(\Delta) = \Sigma(\Delta)$ for all Δ ?
- For a given entailment method Υ and a given Σ, does the method obtain at least the intended models, i.e., is $\Upsilon(\Delta) \supseteq \Sigma(\Delta)$ for all Δ ?
- For a given entailment method Υ and a given Σ, does the method obtain at most the intended models, i.e., is $\Upsilon(\Delta) \subseteq \Sigma(\Delta)$ for all Δ ?

A positive answer to the first question asserts correctness of Υ with respect to Σ. An entailment of all intended models with some not intended ones (the second question) guarantees that the method is *sound* — it may fail to obtain all intended conclusions, but avoids unintended conclusions. And finally, an exclusion of all non-intended models, at the price of possibly missing some intended ones (the third question) ensures that the entailment method is *complete* — it captures all the intended conclusions, together with some unintended.

2.2.3 Selection-equivalence

In the spirit of the systematic methodology we would like to define the selection function $Res(w, e)$ in an abstract way — as a function that specifies *intended* successor states according to fundamental principles. This will allow the agent to maintain a variety of entailment methods. Then, a particular entailment method Υ (i.e., an action theory from a certain class) capturing particular successor states $Res_{\Upsilon}(w, e)$ may be analysed with respect to the successor states $Res(w, e)$.

Following [46], we shall say that an action theory (or a reasoning system based on an action theory) with a function $Res_1(w, e)$ is *selection-equivalent* to an action theory with a function $Res_2(w, e)$ if and only if $Res_1(w, e) = Res_2(w, e)$, for every action e and state w. In other words, action systems using different definitions of selection functions may be inter-translatable. Ideally, in order to have well-defined translations, we need to provide a general unifying semantics for a class of action theories. Then it would become possible to achieve a selection-equivalence between a generic action system based on the abstract selection function $Res(w, e)$, and each one of the action systems of our motivating approaches. Perhaps, more importantly, it would allow us to compare various action systems and gain insight into possible underlying mechanisms.

In this section we argued for the explicit inclusion of the set $\mathcal{D}$ of all legitimate states in our semantical framework, and introduced the function $Res(w, e)$ selecting successor states. A precise definition of this function, ensuring desired selection-equivalence, will be given later. Before we proceed towards a further discussion on the principles of change underlying our general unifying semantics, let us make clear that we have not restricted the selection function $Res(w, e)$ in any way.

First of all, we did not assume that for any action e and state w,

$$Res(w, e) \subseteq [e].$$

In other words, we do not require that a successor state necessarily satisfies the action's direct effects. Although this is a very sensible assumption, it was argued in some proposals that a successor state may be a result merely *triggered* by an action's post-conditions [63]. We will refer to the view that a successor state must satisfy the action's direct effects as *conservative*, and highlight this distinction in further analysis.

In addition, we do not make another intuitively appealing assumption that for any action e,

$$\mathcal{D} \cap [e] \neq \emptyset.$$

This means that we do not impose the requirement that, for every action, there must always be a state where the action post-conditions co-exist with domain constraints. This requirement may, however, become quite reasonable if the conservative view on successor states is accepted. Other possible constraints will be considered later.

2.3 Information States

It would be too unrealistic to assume that an agent's reasoning about changes occurring in the external environment, replicates exactly the external state transitions. It is well known in the reasoning about change research field that one may distinguish two distinct transitions: the transition between the world states r_1 and r_2, brought about by an event e, and the transition between the agent *information* states γ_1 and γ_2, triggered by the perception of the event e or execution of the action e. For example, the distinction has been identified by Peppas [44] who distinguished between a general process called *system dynamics* described as "the process by which a dynamic world changes states due to the occurrence of the events", and another general process called *knowledge dynamics* explained as "the process by which an agent changes beliefs about the current world state, in the light of information about the occurrence of an event".

In other words, system dynamics relates to transitions between world states $\mathcal{W}$, and knowledge dynamics is a cognitive process involving transitions between information states. While this distinction has been identified in [44], the information states were not included in the formal consideration. Instead, the difference was related to different forms of belief change — revision and update. We suggest that, for the purposes of a unifying semantics for reasoning about action and change, a better way to capture this fundamental distinction is to explicitly introduce the finite set of all information states, denoted as Γ.

Intuitively, an information state is a state (or a stage) of an agent's reasoning process. While reasoning, the agent may consider previous and current states of the external world, contemplate potential histories of state transitions, contextualise causal knowledge, and so on. All these rather partial information sources contribute to the reasoning process and fuse into more comprehensive information states. Therefore, in a typical case, an information state has more dimensions than a state of the external world entertained by the agent. Although the term *information state* has undoubtedly a cognitive flavour, we do not intend here to associate this notion with any particular neuro-biological concept, such as conscious, mental, or brain states. What is important is the distinction between knowledge dynamics in the information state-space Γ, and system dynamics in the state-space $\mathcal{W}$.

It is also important to realise that we do not necessarily question the view of logical atomism, or the position that a world state can be completely described by all relevant facts about it, or the agents' ability to reason about both conceivable and nomologically possible world states. Rather, we extend these views by allowing the agent to entertain information states with entirely different dimensions. Most importantly, an information state does not have to be uniquely associated with a time reference. Intuitively, the agent, motivated by a single action (event) e, may imagine a whole series of state transitions in the information state-space Γ before making a judgement on the world transition from $w \in \mathcal{W}$ to one of the successor states in $Res(w, e)$. In short, information states are *not* the agent's beliefs about the state of the world, but *abstract points in the information space that the agent's reasoning may navigate through*.

From a technical point of view, some action domains may avoid the distinction, re-

sulting in equating the world and information state-spaces: $\mathcal{W} = \Gamma$. In other words, each information state entertained by an agent corresponds exactly to one world state. This approximation may well be the reason for fusing the information state-space with the external world state-space in many various approaches. Some recent proposals discern the difference by employing concepts of hyper-states [50] and power-states [51], but without giving a clear intuition on the nature of these additional concepts. We will demonstrate that some action theories may be captured by our semantics while staying with the approximation $\mathcal{W} = \Gamma$, whereas others require $\mathcal{W} \neq \Gamma$.

The inclusion of the information state-space Γ in our framework suggests that we also define some auxiliary concepts. First of all, we introduce a *projection* function $\mathcal{P}$ from Γ to $\mathcal{W}$ — this function maps an information state $\gamma \in \Gamma$ to a world state $w \in \mathcal{W}$. Intuitively, the projection function "extracts" the world state "component" from a more convoluted information state. We require that for any state $w \in \mathcal{W}$ there exists an information state $\gamma \in \Gamma$ such that $\mathcal{P}(\gamma) = w$. In other words, the function $\mathcal{P} : \Gamma \rightarrow \mathcal{W}$ is a surjection (i.e., the function's image is its codomain, and $\mathcal{P}$ can return any value in $\mathcal{W}$).

We also derive a *set-projection* function $\mathcal{X}$ mapping sets of information states onto sets of world states from $\mathcal{W}$. In other words, the function $\mathcal{X} : 2^{\Gamma} \rightarrow 2^{\mathcal{W}}$ is defined as follows: $\mathcal{X}(\{\gamma_1, ..., \gamma_n\}) = \{\mathcal{P}(\gamma_1), ..., \mathcal{P}(\gamma_n)\}$. It is clear that for any set $\Pi \subseteq \Gamma$, we obtain that

$$\mathcal{W} = \mathcal{X}(\Pi) \cup \mathcal{X}(\Gamma \setminus \Pi) = \mathcal{X}(\Gamma),$$

although, in general,

$$\mathcal{X}(\Pi) \cap \mathcal{X}(\Gamma \setminus \Pi) \neq \emptyset.$$

In addition, we define a set of information states $[e]^{\Gamma}$ as $\{\gamma \in \Gamma : \mathcal{P}(\gamma) \in [e]\}$. In other words, $[e]^{\Gamma}$ denotes the set of information states whose world state-space projections are contained in the set $[e]$. By definition, $\mathcal{X}([e]^{\Gamma}) = [e]$.

The following abbreviation will also prove to be useful: $\mathcal{D}^{\Gamma} = \{\gamma \in \Gamma : \mathcal{P}(\gamma) \in \mathcal{D}\}$. The set $\mathcal{D}^{\Gamma}$ contains all the information states that would project to the legitimate states in $\mathcal{D}$. Again, by definition, $\mathcal{X}(\mathcal{D}^{\Gamma}) = \mathcal{D}$. It is worth pointing out that the information states outside $\mathcal{D}^{\Gamma}$ are not illegitimate information states *per se* — the notion of being

legitimate applies only to the states in $\mathcal{W}$. In fact, the information states outside $\mathcal{D}^\Gamma$ are quite acceptable and useful in the reasoning process, providing important intermediate steps for state transitions.

Figure 2.4 illustrates the information state-space Γ (the top part of the figure), while showing the states in $\mathcal{D}^\Gamma$ as filled circles, and enclosing elements of $[e]^\Gamma$ in the polygon. The projection $\mathcal{P}$ of some information states onto the state-space $\mathcal{W}$ (the bottom part of the figure) is shown with arrows.

Clearly, one may derive even more structures on the information-space using the projection function $\mathcal{P}$. However, we will postpone further definitions till technical chapters. At this stage, we just emphasise a new component of our framework — the information-space Γ, together with the projection function $\mathcal{P}$.

Let us also illustrate how these notions can be used in the selection function. Consider again the simple function choosing all legitimate states compatible with an action's direct effects. Now, we can represent this in terms of information states as

$$Res^*(w, e) = \mathcal{X}(\mathcal{D}^\Gamma \cap [e]^\Gamma) = \mathcal{D} \cap [e] = Res_*(w, e).$$

In other words, this function selects projections of those information states that are common to both sets $\mathcal{D}^\Gamma$ and $[e]^\Gamma$.

2.4 Preference Relation on States

One of the problems with the selection function choosing all legitimate states compatible with an action's direct effects is that successor states in $Res^*(w, e)$ are not related to the initial state w. It is hard to imagine that, in general, an action results in a state which may be arbitrarily different form the initial one. As was mentioned in previous sections, one way to address this question is to assume, in some sense, the existence of inertia and apply the Principle of Minimal Change.

There are, at least, two interpretations of this principle. According to the first view, the Principle of Minimal Change is an *intrinsic* property of reality, described by Peppas [44] as follows:

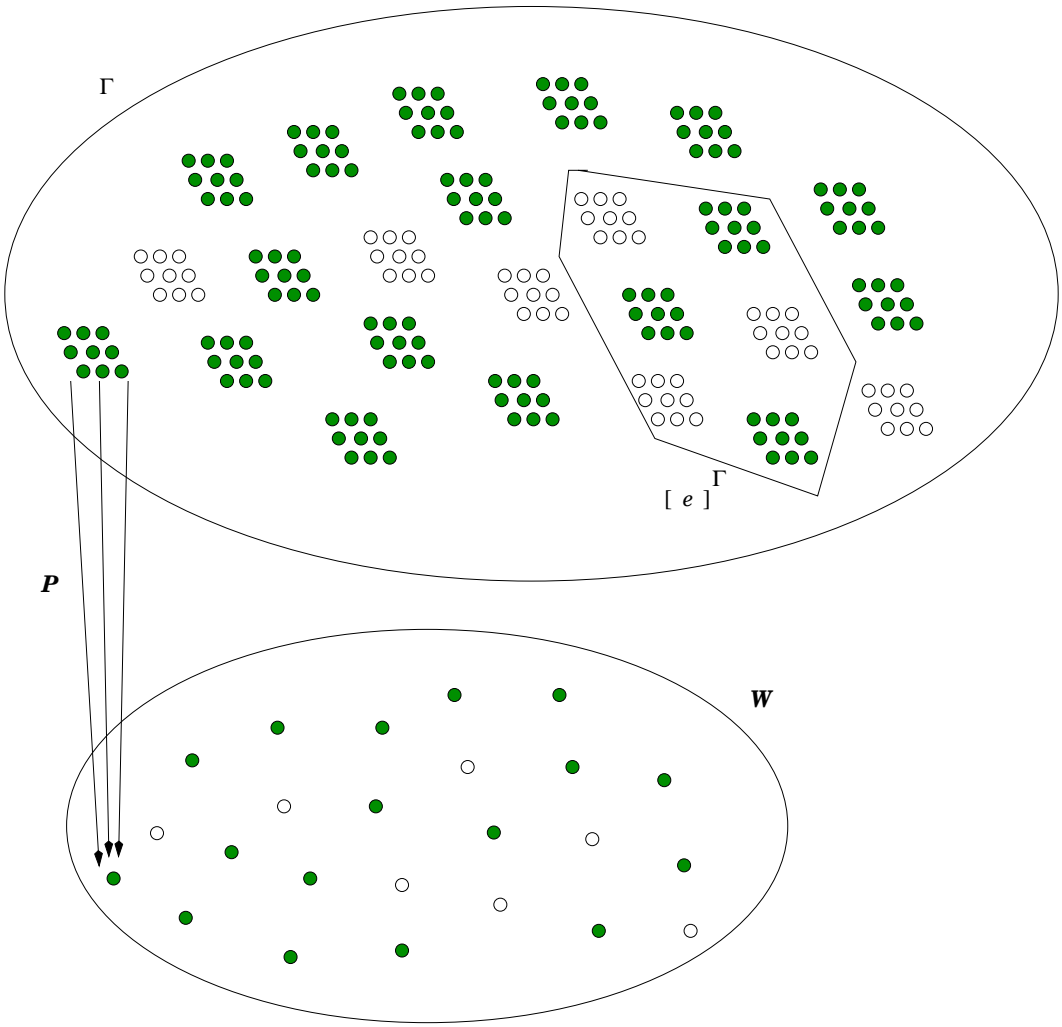

Figure 2.4: The information state-space Γ and the projection function $\mathcal{P}$.

> The universe develops in time in such a way that, its state at any time point is a minimal change of its state at the previous time point, given the entirety of events that have occurred. Minimality is defined with respect to some *global* measure of change.

Moreover, as noted in [44], any dynamic system (world) is an abstraction of only a very small part of the entire universe, and therefore,

> ...in order to establish that minimal change in the universe corresponds to minimal change in a dynamic world, we need to assume that the global measure of change decomposes to *local* measures of change for the different parts of reality, so that a transition between two states of the universe is globally minimal if and only if the transitions between the corresponding parts of the two states are locally minimal.

The second interpretation rejects the intrinsic nature of the Principle, and argues for its introduction during the reasoning process as an *approximation* of the system dynamics. Following Peppas [44], such an approximation should be accurate for the chosen level of abstraction, and, moreover, should provide some structure to world state transitions, facilitating the task of reasoning about action.

The formalisations of dynamic worlds advanced in [44] did not depend on which interpretation is preferred (although, informally, the second reading was favoured). The reason was that in both cases (either an intrinsic ontological property of the world or an epistemic approximation of the world's dynamics) the Principle of Minimal Change reads the same: whenever an event e occurs at some world state w, a successor state $r \in Res(w, e)$ must satisfy the post-condition $[e]$ and differ as little as possible from w with respect to some (local) measure of change. Constructively, there exists an ordering on states $<_w$ reflecting the comparative degree of change between w and a potential successor state. This allows an agent to judge that $r_1 <_w r_2$ if and only if the degree of change between w and r_2 is at least as great[3] as the one between w and r_1. It should be clear that an agent normally maintains a different ordering $<_w$ for each state w, resulting in a set of orderings (a preferential structure).

[3]For simplicity, we shall use the notation $<$ for a non-strict ordering.

The question that naturally arises now is whether we wish to consider a preferential structure only on world states in $\mathcal{W}$ or on information states in Γ as well. This would not, of course, identify minimal change in the world with minimal change in the agent's beliefs about the world[4] — simply because information states are different from the beliefs and may be entitled to their own preferential structure. Besides, a preferential ordering on information states would not introduce syntax-dependent measures of change.

Therefore, it is quite permissible, we believe, to consider a set of orderings on world states or a set of orderings on information states — either method captures the Principle of Minimal Change in its own way. It also appears that one should start with a weaker assumption that only one such set is needed in a general semantics. Normally, it should be possible to derive one preferential structure from another — in other words, a combination of the two may be redundant. More precisely, for a projection function $\mathcal{P}$, and a given specification of an ordering $<_w$ (for each w) defined on $\mathcal{W} \times \mathcal{W}$, one may produce an ordering $\ll_\gamma$ (for each γ) defined on $\Gamma \times \Gamma$, and vice versa. Of course, future developments may highlight a possibility that two independent preferential structures are required — one for each of the state-spaces.

In this work, we choose to operate with a preferential structure $\mathcal{O} = \{<_\gamma\colon \gamma \in \Gamma\}$ defined on information states, without making the claim that it is richer or more intuitive than a structure on world-states. In fact, when we discuss selection-equivalence with an action theory that needs the distinction $\mathcal{W} \neq \Gamma$ (Chapter 5), we derive preferential orderings for information states $\ll_\gamma$ out of simple orderings $<_w$ defined for world states. In short, the inclusion of a preferential structure on information states in our framework is a matter of choice, and one can obtain equivalent characterisations with a preferential structure on world states.

Before we apply the new concept to construction of some selection functions, let us briefly discuss some natural properties of an ordering $<_r$, regardless of its domain. The following properties seem to be intuitive, and were suggested in the past:

($\mathcal{O}_1$) Reflexivity: $\quad p <_r p$.

[4]This was, in fact, the main reason for a failure of the Possible Worlds Approach [17], as shown by Winslett [70].

($\mathcal{O}_2$) Transitivity: if $p <_r q$ and $q <_r x$ then $p <_r x$.

($\mathcal{O}_3$) Discreteness: r is the single minimal element in its state-space with respect to $<_r$.

($\mathcal{O}_4$) Minimality: every non-empty subset of the state-space has a minimal element with respect to $<_r$.

The reflexivity condition is straightforward. Transitivity also seems to be very intuitive: if the degree of change between r and p is no greater than the degree of change between r and q (represented as $p <_r q$), and the latter is no greater than the degree of change between r and x (represented as $q <_r x$), then it is quite safe to assume that a change from r to p is no greater than a change from r to x (represented as $p <_r x$).

The last two conditions make explicit use of minimality with respect to an ordering $<_r$. This minimality is defined in the usual way: given an ordering $<_r$, a state p is $<_r$-minimal in a subset A of its state-space if and only if there is no other element $q \in A$ such that $q <_r p$.

In general, we may define a set $min(<_r, A)$ as a subset of A containing states nearest to the state r in terms of the ordering $<_r$. In other words, $min(<_r, A) = \{p \in A : \neg\exists q \in A,\ q \neq p,\ q <_r p\}$. Then, any element of $min(<_r, A)$ may be referred to as a state $<_r$-minimal in A.

Now, the third condition specifying that any state r is the single minimal element with respect to an ordering $<_r$ centered on itself has the intuitive justification that no state is "more similar" to r than r itself. It is also related to the assumption that change is discrete, or in other words, that there are no state transitions with an infinitesimally small change: $\ldots <_r r_3 <_r r_2 <_r r_1$ [44]. This assumption, in fact, allows us to justify the last condition on minimality as well — every non-empty subset of the state-space has a minimal element with respect to each $<_r$.

These conditions imply, in particular, that an ordering $<_r$ is an inductive partial preorder with r as its first element [44].

One interesting ordering type satisfying the conditions ($\mathcal{O}_1$) - ($\mathcal{O}_4$) is the so-called PMA ordering, based on the Possible Models Approach [70]. In order to describe this ordering constructively, we need to assume certain internal structure for states. For instance, let us consider n basic truth-valued fluents, and let each state be a set with n elements such that each of the basic features or its negation appears as an element. We also define the symmetric difference between two states x and y to be the set $\mathit{Diff}(x, y) = (x \setminus y) \cup (y \setminus x)$, where $x \setminus y$ denotes set subtraction. For example, if $r = \{a, b, c\}$, $p = \{a, \neg b, c\}$, and $q = \{a, \neg b, \neg c\}$, we obtain $\mathit{Diff}(r, p) = \{b, \neg b\}$ and $\mathit{Diff}(r, q) = \{b, c, \neg b, \neg c\}$.

Now, we shall say that a state y is preferred to a state z relative to x in terms of the PMA ordering $\prec_x$, denoted $y \prec_x z$, if and only if $\mathit{Diff}(x, y) \subseteq \mathit{Diff}(x, z)$. Intuitively, it means that state y differs less from x than the state z does from x in terms of basic features. Continuing the example with three states r, p and q, we immediately obtain that $p \prec_r q$. Obviously, the PMA ordering is not a total order. For example, given a state $s = \{a, b, \neg c\}$, we cannot say if $s \prec_r p$ or $p \prec_r s$, because neither of the symmetric differences $\mathit{Diff}(r, p) = \{b, \neg b\}$ or $\mathit{Diff}(r, s) = \{c, \neg c\}$ is contained in each other. Figure 2.5 depicts direct PMA preferences for 8 states definable with 3 truth-valued fluents a, b and c, with respect to the state $r = \{a, b, c\}$ (more distant states appear further to the right from r). Note that states in each vertical layer are not mutually comparable in terms of the PMA ordering — in other words, PMA ordering is not total. Figure 2.6 shows a subset of the original ordering, such that states from different areas enclosed by pseudo-concentric dashed curves are PMA-comparable.

Having introduced the concept of a preferential structure $\mathcal{O} = \{<_\gamma : \gamma \in \Gamma\}$ defined on information states, we may now discuss a preferential semantics in general. One of the important developments in the field of Reasoning about Action was Shoham's *preferential semantics* [60] for a class of non-monotonic logics. Under this idea an ordering is placed over the class of interpretations. The models corresponding to a particular inference are then identified as the minimal models under this ordering that satisfy the premises. In an intuitive sense, the ordering represents a preference over interpretations with only the most preferred (most plausible) being tolerated as serious possibilities.

In terms of our framework, the preferential structure $\mathcal{O}$ suggests a very intuitive way

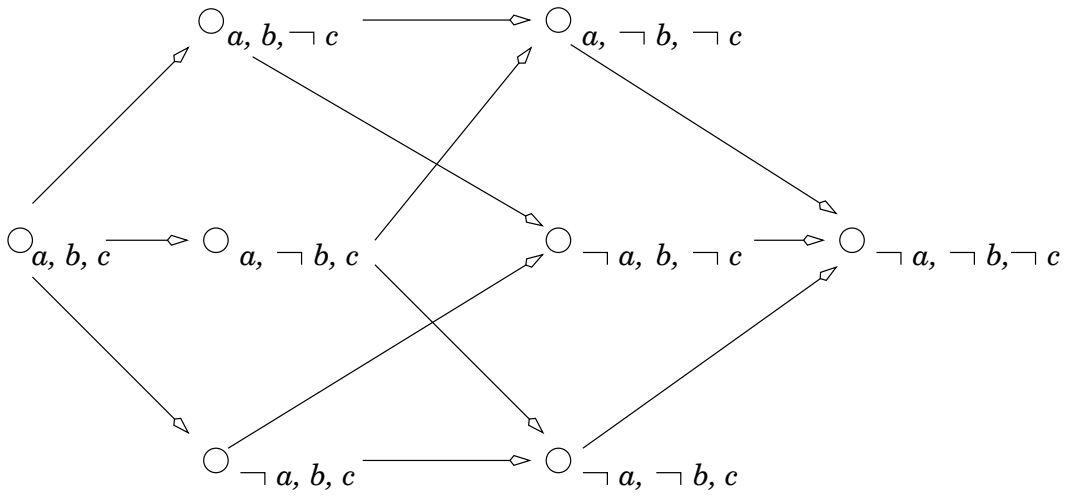

Figure 2.5: A partial PMA ordering: arrows point to more distant states.

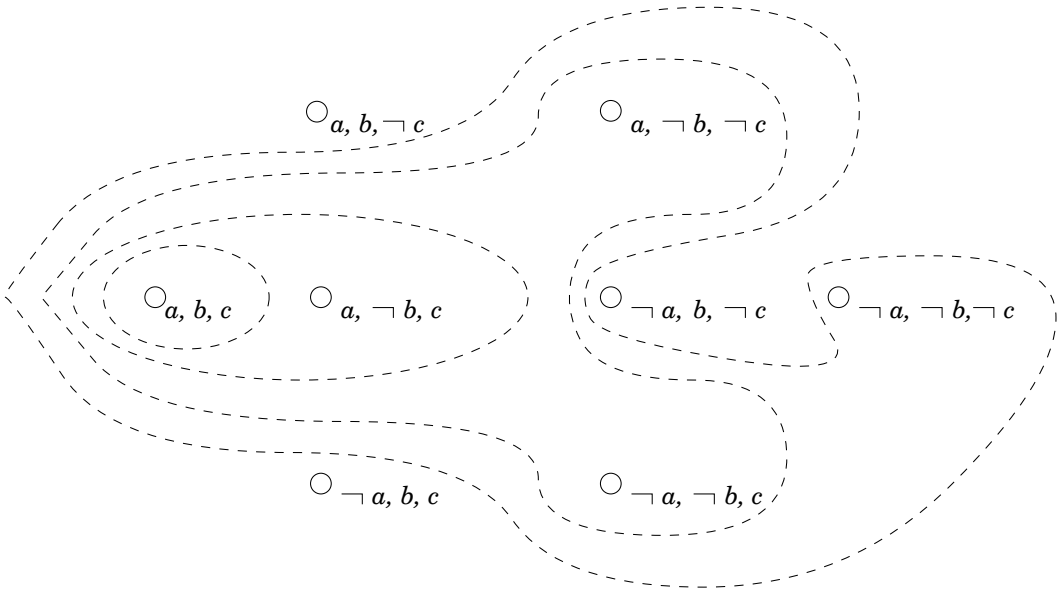

Figure 2.6: A subset of PMA ordering: “expanding” boundaries of minimality.

to define a selection function $Res(w, e)$. Specifically, for a simple case approximating $\mathcal{W} = \Gamma$, the function $Res(w, e)$ should choose those legitimate states compatible with the post-condition $[e]$ that are nearest to w in terms of $<_w$. More precisely,

$$Res_1(w, e) = min(<_w, \mathcal{D} \cap [e]).$$

Figure 2.7 illustrates this selection (again, more distant states, enclosed in the expanding minimality-driven circles, appear further to the right from the left-most initial state w).

We should point out that there exists a stronger version of the preferential selection function:

$$Res_0(w, e) = min(<_w, [e]) \cap \mathcal{D}.$$

Here, an agent first selects the $<_w$-minimal states among the post-condition states $[e]$, and then chooses legitimate states out of the selection, if there are any (Figure 2.7). If, for example, the state y was legitimate, then it would be (uniquely) selected by $Res_0(w, e)$ as well as $Res_1(w, e)$. The function $Res_0(w, e)$ may result in no selection, while the condition ($\mathcal{O}_4$) ensures that $Res_1(w, e)$ always selects at least one state, if the set $\mathcal{D} \cap [e]$ is non-empty; otherwise it also selects no states. Clearly, for any state w and action e,

$$Res_0(w, e) \subseteq Res_1(w, e) \subseteq Res_*(w, e).$$

Informally, we may refer to the selection function $Res_0(w, e)$ as the *first boundary*, and $Res_1(w, e)$ as the *second boundary* of successor states.

We assumed here an approximation $\mathcal{W} = \Gamma$, and used a preferential structure over world-states. Similar "boundary"-setting preferential selection functions can be defined for the case when $\mathcal{W} \neq \Gamma$. Let an information state $\alpha(w)$ be such that $\mathcal{P}(\alpha(w)) = w$ for a state $w \in \mathcal{W}$. Then we may specify the following selection functions.

$$Res^1(w, e) = \mathcal{X}(min(<_{\alpha(w)}, \mathcal{D}^\Gamma \cap [e]^\Gamma)).$$

and

$$Res^0(w, e) = \mathcal{X}(min(<_{\alpha(w)}, [e]^\Gamma) \cap \mathcal{D}^\Gamma).$$

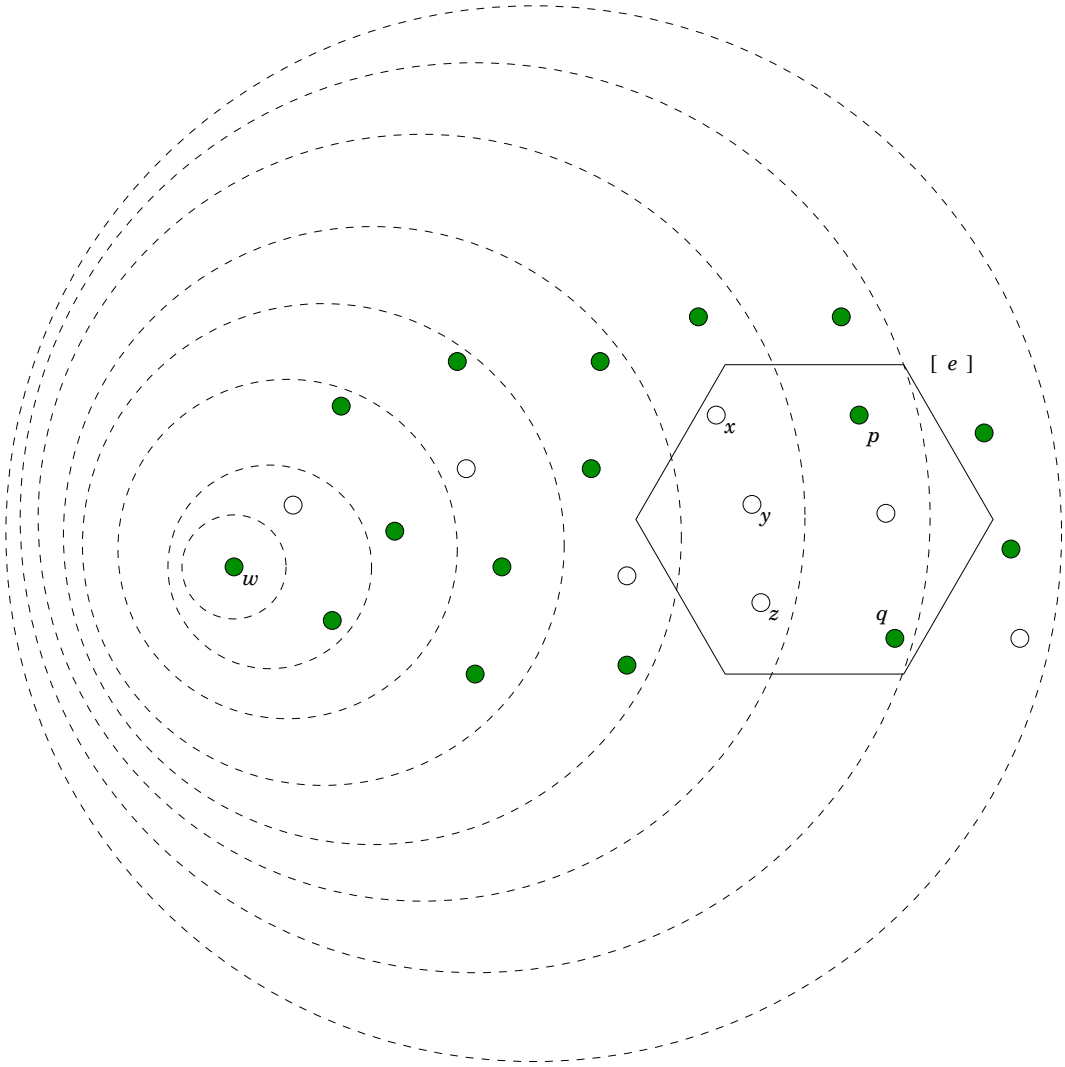

Figure 2.7: The selection function $Res_1(w, e)$ contains two states p and q, while the selection function $Res_0(w, e) = \emptyset$, given that the set $min(<_w, [e]) = \{x, y, z\}$ contains no legitimate states.

Here, the agent selects minimal states in the information state-space by using the preferential structure $\mathcal{O}$ on information states, and then projects them onto successor world states. The function $Res^1(w, e)$ is analogous to the function $Res_1(w, e)$ — it selects the information states that are closest to some state $\alpha(w)$ among the information states whose projections are both legitimate and consistent with the action's direct effects, and then projects the selections to the world state-space $\mathcal{W}$. The function $Res^0(w, e)$ is analogous to the function $Res_0(w, e)$ — it selects the information states that are closest to some state $\alpha(w)$ among the information states whose projections are consistent with the action's direct effects, then chooses only those selections that project onto legitimate world states, and finally projects the remaining selections to the world state-space $\mathcal{W}$.

Figure 2.8 illustrates the expanding minimality layers with dashed curves. It does not matter which of the information states $\alpha(w)$ such that $\mathcal{P}(\alpha(w)) = w$ is chosen to fix the preference ordering $<_{\alpha(w)}$. Again, it is obvious that, for any state w and action e,

$$Res^0(w, e) \subseteq Res^1(w, e) \subseteq Res^*(w, e).$$

A minimisation in the information state-space Γ followed by a projection onto world state-space $\mathcal{W}$ may obviously result in different outcomes compared with a direct minimisation in $\mathcal{W}$. For example, let w, s and q be the only legitimate world states, and α, β and γ be information states such that $\mathcal{P}(\alpha) = w$, $\mathcal{P}(\beta) = s$ and $\mathcal{P}(\gamma) = q$. Let us also assume that $\neg(s <_w q)$ and $\neg(q <_w s)$, in other words, neither of the states s and q is closer to w than the other. For simplicity, let $\beta \ll_\alpha \gamma$ be the only preference in the information state-space with respect to α. Consider now an action e such that only s and q satisfy its post-condition and belong to the set $[e]$. Then $Res_1(w, e) = \{s, q\}$, so that both s and q are selected as successor states, being $<_w$-minimal legitimate states in $[e]$. However, the only $\ll_\alpha$-minimal state in $\mathcal{D}^\Gamma \cap [e]^\Gamma$ is β, ruling out state γ. Therefore, $Res^1(w, e) = \{s\}$, selecting only one successor state s (a projection of β), and leaving out q. Of course, other examples may show that $Res^1(w, e)$ chooses more successor states than $Res_1(w, e)$. In short, neither domain of minimisation ($\mathcal{W}$ or Γ) leads to stronger successor selections in general.

Another important aspect is that, regardless of the domain of the preferential struc-

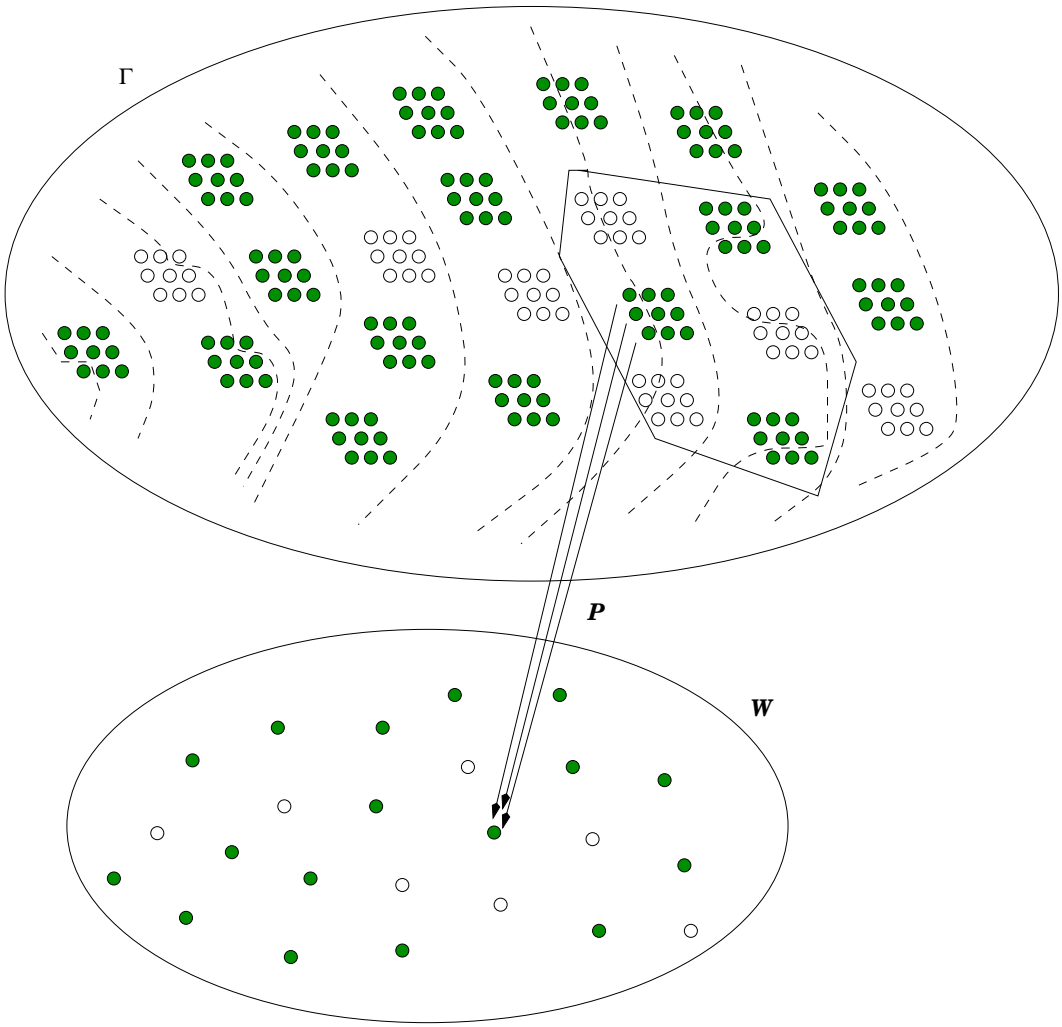

Figure 2.8: Minimisation in the information state-space and projection.

ture, each ordering $<_w$ is dependent only on the initial state w, and is not contingent on an action. In other words, we do not wish to deal with cumbersome preferential structures containing orderings for each state-action pair $<_{w,e}$ — clearly this would seriously undermine the quest for conciseness, given the number of potential combinations and weak *elaboration tolerance*.

In summary, the steps we have taken so far towards a general semantics, can be reflected in a tuple $\langle \mathcal{W},\ \mathcal{D},\ \Gamma,\ \mathcal{E},\ \mathcal{O} \rangle$ and the functions $[e] : \mathcal{E} \to 2^{\mathcal{W}}$ (post-conditions), $\mathcal{P}(\gamma) : \Gamma \to \mathcal{W}$ (projection), and $Res(w, e) : \mathcal{W} \times \mathcal{E} \to 2^{\mathcal{W}}$ (selection).

2.5 Causal Rules, Causal Relations, and Causal Laws

Despite many intuitive features of preferential style semantics, it has been shown recently that sometimes there are additional forces "producing" successor states. The main explanation suggested in the literature so far is that domain constrains (reflected in the legitimate states $\mathcal{D}$) are not sufficient for capturing all the domain dependencies. Therefore, it is typically argued, one needs to introduce (in some form) $causal$ constraints and use the force of causality in complementing minimality. For example, McCain and Turner's causal theory of action [37] uses "causal laws", Thielscher's [63] framework is based on "causal relationships", and Sandewall proposed a causal relation on states in order to capture ramifications cascading to successor states.

The inadequacy of logical domain constraints has been illustrated on many occasions, and we do not intend to elaborate on this. The Light Detector example will be described in Chapter 5 devoted to the causal relationship approach of Thielscher[63], and Chapter 3 on action languages will relate to this issue as well.

Here, however, we intend to discuss some technical aspects of causally driven state transitions. In order to explain what exactly is meant by "causally driven state transitions", we would like to briefly review some related terminology, introduced recently in various logics of action dealing with causality.

2.5.1 Causal Rules — Ontological Dimension of Causation

Two notions — *action-triggered* and *fluent-triggered* causation — were introduced by Lin [33]. It was argued that since normal state (domain) constraints referring to only the truth values of fluents are not strong enough to properly trace the ramifications of actions, "a notion of causation needs to be employed explicitly" [33]. Technically, Lin introduced a new ternary predicate C$aused(p, v, s)$ into the Situation Calculus: C$aused(p, v, s)$ is true if the proposition p is caused by something unspecified to have the truth value v in the state s. Then, using this predicate, two types of causal statement were represented:

- *action-triggered* causal statements (such as that the action $load$ causes the gun to be loaded);
- *fluent-triggered* causal statements (such as that the fact that the $switch$ is in the up position causes the $light$ to be on).

Lin has convincingly argued that *action-triggered* causation is convenient for representing direct effects of actions, and *fluent-triggered* causation — for indirect effects, or ramifications.

Similar types of causality are used in so-called *action languages* — see, for example, [67, 31]. We shall focus on the syntax and semantics of some of these languages in further chapter(s) but, at this stage, we just observe that, typically, an action language includes effect propositions of the form

$$A \textbf{ causes } \varphi \textbf{ if } \psi,$$

where A is an action, and φ and ψ are fluent formulae. In particular, the action description language DL_{if} refers to these propositions as *dynamic causal laws* [31] expressing "event causality" — they tell us which changes are caused by performing an action. The second type of DL_{if} propositions are called *static causal laws*, and capture "fact causality" — a dependency between two facts contained in the same state. A static causal law is represented in the DL_{if} language in the form

$$\varphi \textbf{ causes}_f \psi.$$

Another action language dealing with causality, $\mathcal{AC}$, represents static causal laws with so-called sufficiency propositions [67]

$$\varphi \textbf{ suffices for } \psi.$$

The static causal laws (fact causality) of DL_{if} and sufficiency propositions of $\mathcal{AC}$ share the same semantics ensuring that "one can make the fluent formula ψ true by making the fluent formula φ true" [67]. The original syntax and semantics of (static) causal laws was described in [37], where a proposition

$$\phi \Rightarrow \psi$$

was used to express a causal "determination relation" between ϕ and ψ, working as an inference rule. This original approach and the $\mathcal{AC}$ language do not explicitly refer to *dynamic causal laws*, but specify instead the post-conditions and effect propositions of actions respectively.

In short, the *dynamic* and *static* causal laws capturing event and fact causality in action languages match two forms of causation represented in Lin's framework: *action-triggered* and *fluent-triggered*.

Another similar classification is given by Geffner [14]. The Geffner approach introduces causal rules

$$F \rightarrow A,$$

where F is a formula and A is an atom, that express mechanisms by which the truth of F *normally* causes the truth of A. This time, there is a distinction between *non-temporal* causal rules $F \rightarrow A$ and *temporal* causal rules written as (note the longer arrow)

$$F \longrightarrow A.$$

The difference is that the non-temporal rule $F \rightarrow A$ says that A is normally true in the state s when F is true in s, while the temporal rule $F \longrightarrow A$ says that A is normally true in the state *following* s when F is true in s. Both rule types (temporal and non-temporal) have a strict form, when the mechanism of causing the truth of A "normally" is replaced by one which "always" causes the truth of A. Strict causal rules are denoted as follows:

$$F \supset A,$$
$$F \Longrightarrow A,$$

where the first one is a strict non-temporal rule, and the second — a strict temporal rule.

It is not hard to observe direct parallels between the temporal and non-temporal causal rules of Geffner with the dynamic and static causal laws of action languages, and in turn with the action-triggered and fluent triggered causation of Lin. In addition, Geffner points out that temporal and non-temporal causal rules play the role of *effect axioms* and *ramification constraints* respectively — the latter pair was introduced by Lin and Reiter [32], and later subsumed by action- and fluent-triggered causation.

Following these and other recent approaches to representing causal information, we accept the distinction between action-triggered and fluent-triggered causality. This is not an over-simplification even from a philosophic viewpoint. For example, two sorts of causation subjects are identified by Mellor [41]: *states of affairs* ("sentences, statements or propositions"), and *particulars* (things or events). In other words, the causal connections among states of affairs would correspond to fluent-triggered causality, and the causal connections between particulars are similar to action-triggered causality.

2.5.2 Causal Relation — Epistemological Dimension of Causation

The distinction discussed in the previous section is, obviously, related to the difference between an action's direct effects (action-triggered) and ramifications driven by state constraints (fluent-triggered).

Hence, the attempts of preferential style selection functions like $Res_1(w, e)$ and $Res^1(w, e)$ (defined in Section 2.4) are not entirely adequate, because minimal legitimate states among post-condition states $[e]$ do not necessarily reflect fluent-triggered causality. In other words, we believe that the post-condition function $[e]$ together with a preferential structure on Γ captures in most cases only the aspect of action-triggered causality. The intuitive reason for this is simple: while all the states in $[e]$ agree on the post-conditions of the action in question, the selection of minimal states may often be insufficient for capturing various causal dependencies.

It is not entirely inconceivable that a more intricate preferential structure defined on

a high-dimension information state-space will solve the "puzzle" and completely characterise successor states, without employing causation. However, it is quite plausible that such an approach may require some non-obvious assumptions, or severely restrict classes of action theories. For that reason, we are inclined to investigate an approach, where both Principles of Minimal and Causal Change are given clear roles. Even if some of the examples used to justify an explicit introduction of causal dependencies in action theories are refuted in the future, we believe that a concise semantics explicitly embedding causality may cover a broader class of action theories.

Therefore, we intend to augment our framework with a component explicitly targeting fluent-triggered causality, while retaining existing structures. One immediate difficulty is that fluent-triggered (fact) causality is not expressed in extensional terms (it directly refers to the internal structure of states), at least in the examples we considered in this section. More precisely, the examples used a causal relation defined on fluents or basic "states of affairs".

However, there is a simple solution. Instead of committing to a formal logical language and specifying fluent-triggered causal statements in some syntactic form, it is possible to capture the underlying constraints in a causal binary relation on states. This choice manifests another dual aspect of causation — it may be expressed both intensionally (with a causal relation on fluents) and extensionally (with a causal relation on states). This duality is "orthogonal" to the event-fact distinction, in the sense that both types of causation (event driven and fact driven) may be represented with and without references to the internal structure of states (Figure 2.9). For example, the causal relation *causes* can refer to (internal) state variables and/or events, while the causal relation "$->$" can be (extensionally) defined on sets of states. Here, the notation $X -> X \cap Y$ may indicate that the agent's reasoning process propagates from the states in the set X to the states in the set $X \cap Y$.

Consequently, we intend to introduce a binary relation $\mathcal{M}$ on information states, in order to capture fluent-triggered causality extensionally. Two aspects require some explanation: the choice of the domain of the relation $\mathcal{M}$, and its properties.

The choice of the information state-space Γ (and not the state-space $\mathcal{W}$) as the domain for the relation $\mathcal{M}$ is not arbitrary. Technically, it extends our options in expressing

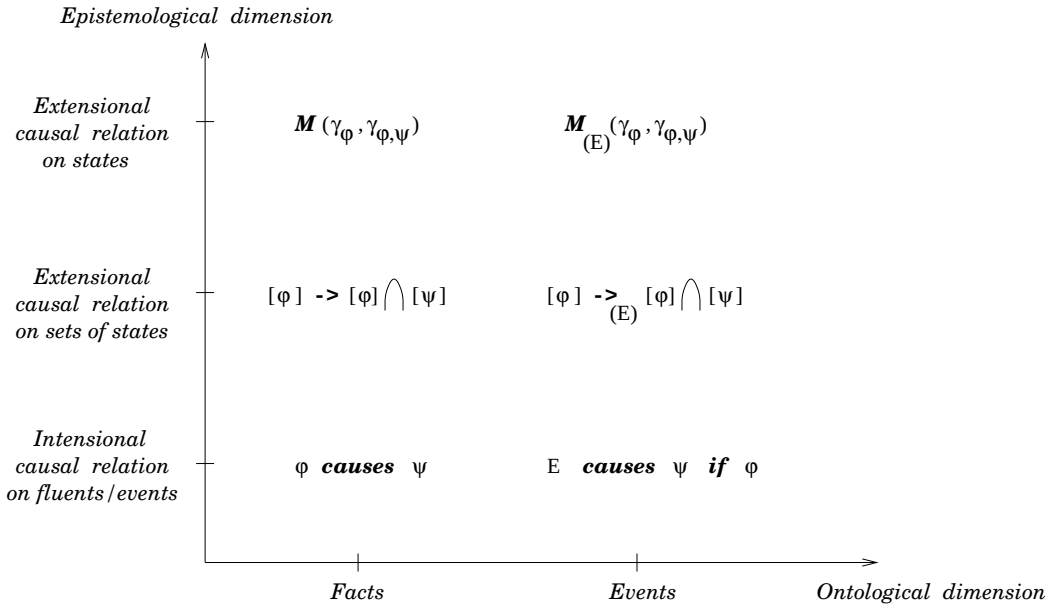

Figure 2.9: Ontological and epistemological dimensions of causation. Causal relations can be defined on states of affairs (fluents, events), on sets of states, on (information) states, etc.

a variety of state to state transitions driven by causation because, typically, the state-space Γ has a higher dimension than the state-space $\mathcal{W}$. We may need this extra dimensionality in capturing some not so obvious causal connections.

It is quite important to realise that the choice of Γ as the domain for the relation $\mathcal{M}$ does not commit us to a particular stand on the nature of causation in terms of its status as a category. Is causation an *ontological* category? Is it a purely *epistemological* category, a theoretical relation, which belongs exclusively to our account of experience? A positive answer to the latter question, taken by empiricism, argues that "the status of causation category is purely epistemological, that is, causation concerns solely our experience with and knowledge of things, without being a trait of the things themselves" [5, p. 5]. According to Bunge [5, p. 6] who criticised the empirical doctrine,

> ...causation is not a category of *relation* among *ideas*, but a *category of connection and determination* corresponding to an actual trait of the factual (external and internal) world, so that it has an ontological status — although,

> like every other ontological category, it raises epistemological problems. Causation, as here understood, is not only a component of experience but also an objective form of interdependence obtaining, though only approximately, among real events, i.e., among happenings in nature and society.

The reason why the choice of Γ and not $\mathcal{W}$ as the domain for the relation $\mathcal{M}$ does not commit us to either view, is the following. If one takes the view advocated by Bunge and considers an objective form of interdependence $\Rightarrow$ between two real states of affairs (or events) φ and ψ, then "*there can be no causal links among states*" [5, p. 71]. As pointed out by Bunge, the state of a dynamic material system is a system of qualities, not an event or string of events: "every state is the *outcome* of a set of determiners ...", and consequently, "there can be no action of one state upon another state of a given system". This argument would rule out the state-space $\mathcal{W}$ as the domain for the causal relation $\mathcal{M}$.

At the same time, the argument that causation is "an objective form of interdependence obtaining ... among real events", admits that causation may manifest itself "only approximately". Therefore, the agent may need to fill the gap between approximated causal constraints and intended ramifications. Hence, defining the causal relation $\mathcal{M}$ in terms of information states allows us to better approximate instances of interdependence between "real events" [5]. In other words, we believe that it is fairly permissible to express the interdependence $\Rightarrow$ between two real states of affairs (or events) φ and ψ via a relation $\mathcal{M}_{(\Rightarrow)}$ between some relevant information states γ_φ and $\gamma_{\varphi,\psi}$.

The same argument (approximation of causation) would, of course, apply if one follows the empirical doctrine. In this case, however, one would probably be more satisfied with "upgrading" the *theoretical* causal relation to the information state-space, and leaving functional, quantum and other objective dependencies to the world state-space. Thus, if causation does not apply to things but to experience alone, and is nothing but a direc-

[5] The view that causation is an an objective form of interdependence is, of course, related to the question of causal asymmetry and "time's arrow". As mentioned, for example, in [47], "[p]hysicists in particular have been interested in the question as to whether there is a single 'master arrow', from which all the others are in some sense derived ... the leading candidate for this position has been the so-called arrow of thermodynamics. This is the asymmetry embodied in the second law of thermodynamics, which says roughly that the entropy of an isolated physical system never decreases."

tion enabling us to order or to label phenomena, then there is no harm in defining it in the information state-space.

So, in short, it would be incorrect to say that favouring Γ over $\mathcal{W}$ brings us closer to the empiricist view on causation[6].

Now let us discuss the second aspect — properties of the relation $\mathcal{M}$. First of all, let us quote from Bunge [5, p. 240] who argues that the causal problem is an ontological question, that

> ...like every other philosophic and scientific question, the causal problem raises epistemological problems ...and logical problems. The logical side of the causal problem essentially consists in the logical structure of propositions by means of which causal statements are formulated.

Bunge pursues this point with a clarification [5, p. 244] that

> ...what is required is not an extension of formal (extensional) relations, but the determination of a type of semantic connection among terms that are relevant to each other: *the logical aspect of the causal problem is semantical rather than syntactical*.

To some extent following Bunge, we attempt to specify "the topology of the series, not the nature of its terms", while taking weakest assumptions on formal properties of the binary relation $\mathcal{M}$ holding among elements interpretable as information states:

($\mathcal{M}_1$) Irreflexivity: $\neg\mathcal{M}(\alpha,\alpha)$.

($\mathcal{M}_2$) Asymmetry: if $\mathcal{M}(\alpha,\beta)$ then $\neg\mathcal{M}(\beta,\alpha)$.

The first property is straightforward: *nihil est causa sui* and, together with the last one, asserts the directionality of causation. There is another property, suggested in the literature:

[6] If one wanted to illustrate that causation is a *purely* ontological category, then the intensional causal relation on fluents/events would have to be placed right on the horizontal (ontological) axis in Figure 2.9.

($\mathcal{M}_3$) Transitivity: if $\mathcal{M}(\alpha, \beta)$ and $\mathcal{M}(\beta, \gamma)$ then $\mathcal{M}(\alpha, \gamma)$.

Transitivity, however, is a more debatable property than $\mathcal{M}_1$ and $\mathcal{M}_2$. We prefer not to require it and use, instead, the transitive closure of the relation $\mathcal{M}$, denoted $\mathcal{M}^*$, when needed. In other words, we rely on the propagation in the information state-space driven by the relation $\mathcal{M}$ without postulating the "cause-effect" relations between states not related in $\mathcal{M}$.

The binary relation $\mathcal{M}$ defined on $\Gamma \times \Gamma$ creates a multiplicity of causal chains $\mathcal{M}(\gamma_1, \gamma_2), \mathcal{M}(\gamma_2, \gamma_3), ..., \mathcal{M}(\gamma_i, \gamma_{i+1}), ..., \mathcal{M}(\gamma_{k-1}, \gamma_k)$, and therefore enables a propagation in the information state-space along them. Intuitively, in terms of reasoning about action, such a transition process propagates some initial *change* towards an information state γ_k that is *stable* in terms of $\mathcal{M}$. In other words, the propagation stops when there are no possible causal links from γ_k. In short, the most essential role of the binary relation $\mathcal{M}$ is to provide some "topology" for tracing causal ramifications in Γ, in addition to ruling out non-stable states.

Let us denote by $\mathcal{K}_\mathcal{M}$ the set of stable information states $\{p \in \Gamma : \neg\exists q \in \Gamma, \mathcal{M}(p, q)\}$. We also require

($\mathcal{M}_\mathcal{D}$) Density: $\mathcal{D} \cap \mathcal{X}(\Gamma \setminus \mathcal{K}_\mathcal{M}) = \emptyset$,

In other words, no unstable information state may be projected onto a legitimate state. The density condition implies

$$\mathcal{D} \subseteq \mathcal{X}(\mathcal{K}_\mathcal{M}).$$

Therefore, although a stable information state may be projected onto an illegitimate world state, a legitimate state is always a projection of a stable information state. This restriction reflects a possibility that some domain constraints may eliminate more illegitimate states than the causal relation alone. It could be argued that domain constraints (reflected in the set $\mathcal{D}$) correspond to non-causal (functional, etc.) laws, while the elements of $\mathcal{X}(\mathcal{K}_\mathcal{M})$ are just those states that do not conflict with causation in a given

domain, from the agent's point of view. Obviously, if $\mathcal{W} = \Gamma$, the original specified density condition reduces to

$$\mathcal{D} \subseteq \mathcal{K}_{\mathcal{M}}.$$

2.5.3 Causal Laws — Nomological Dimension of Causation

In this section we briefly consider the fundamental issues of the existence of underlying causal laws and their possible connections with causal relation(s). These questions may be raised regardless of whether the causal relation refers to the internal structure or ignores it, and whether causation is action-triggered or fluent-triggered. In other words, interdependence between causal relations and laws highlights another dimension of causation — namely, the nomological dimension (Figure 2.10).

Let us introduce this problem with a succinct quotation from Tooley [65, p. 252]:

> There are a number of causal concepts. Some are concepts of relations between states of affairs (or events). One state of affairs causes another. Or it is causally sufficient in the circumstances for another. Or it is causally necessary in the circumstances for the other. Or one states of affairs is, by itself, causally completely sufficient for another. Or it is causally completely necessary for the other.
>
> Other causal concepts are of relations between, not states of affairs, but $types$ of state of affairs. Thus, for example, one type of state of affairs may be causally sufficient to ensure the existence of a state of affairs of some other type. Or it may be causally necessary for the existence of a state of affairs of some other type.

It should not be a great over-simplification then, to identify causal relations with token-causality and causal laws with type-causality, where the latter captures certain sorts of *regularity* or *recurrence*. The question then may be formulated as whether all specific (individual) causal relationships are instances of general (universal) causal laws. However, as noted by Brother William (in Umberto Eco's "The Name of the Rose" [9, p. 206]),

> . . . if only the sense of the individual is just, the proposition that identical causes have identical effects is difficult to prove. A single body can be cold or hot, sweet or bitter, wet or dry, in one place — and not in another place. How can I discover the universal bond that orders all things if I cannot lift a finger without creating an infinity of new entities? For with such a movement all the relations of position between my finger and all other objects change. The relations are the ways in which my mind perceives the connections between single entities, but what is the guarantee that this is universal and stable?

Indeed, this topic "has sown doubts" not only in the learned Franciscan's mind, but also in minds of many prominent philosophers.

According to one (reductionist) view, causal laws are primary, and causal relations are secondary. This view, shared by a majority of philosophers, can be expressed as *the thesis of the Humean supervenience of causal relations* [65, p. 29]:

> The truth values of all singular causal statements are logically determined by the truth values of statements of causal laws, together with the truth values of non-causal statements about particulars.

If this view is correct, then the fundamental concept is that of a causal law. Consider two possible worlds that agree on all causal laws, and on all non-causal properties of, and relations between, particular states of affairs or events. Then, according to the reductionist position, these two worlds must also agree on all causal relations between states of affairs or events.

An opposite (singularist) viewpoint, suggesting that causal relations between states of affairs are primary, and causal laws are secondary, has also been taken. C. J. Ducasse [8, p. 129], for example, persuasively argued that

> The supposition of recurrence is thus wholly irrelevant to the meaning of cause; that supposition is relevant only to the meaning of law. And recurrence becomes related at all to causation only when a law is considered which happens to be a generalization of facts themselves individually causal

> to begin with. A general proposition concerning such facts is, indeed, a causal law, but it is not causal because general. It is general, i.e. a law, only because it is about a class of resembling facts; and it is causal only because each of them already happens to be a causal fact individually and in its own right (instead of, as Hume would have it, by right of its co-membership with others in a class of pairs of successive events). The causal relation is essentially a relation between concrete individual events; and it is only so far as these events exhibit likeness to others, and can therefore be grouped with them into kinds, that it is possible to pass from individual causal facts to causal laws.

According to an extreme singularist view, it is possible for two events or states of affairs to be causally related without that relation being an instance of any law.

This position allowed C. J. Ducasse to advocate, in particular, that causal connection is not an objective connection, and "is not a sensation at all, but a relation ... which has individual concrete events for its terms" [8, p. 132]. However, it is quite clear that a particular nomological stand should not commit one to a specific point in the ontological-epistemological plane. In other words, it is not inconceivable that causation is an ontological connection, held (in some domains) primarily between states of affairs, while causal laws are secondary[7]. In short, the nomological aspect of causation is orthogonal to the ontological-epistemological plane, and generates another dimension in the multi-faceted causation space.

One way of reconciling the seemingly polar reductionist and singularist views was suggested by Davidson [7, p. 85], who made a distinction between "knowing there is a law 'covering' two events and knowing what the law is: in my view, Ducasse is right that singular causal statements entail no law; Hume is right that they entail there is a law".

As an aside, we should mention another interesting alternative to the reductionist (supervenience) view, on the one hand, and the singularist view on the other, proposed by Tooley as a realist position [65, p. 175]:

[7] Although, then "it must be logically possible to have causal relations for which there are no covering laws. Or even for there to be a world that is full of causally related events, but which is completely anomic" [65, p. 175].

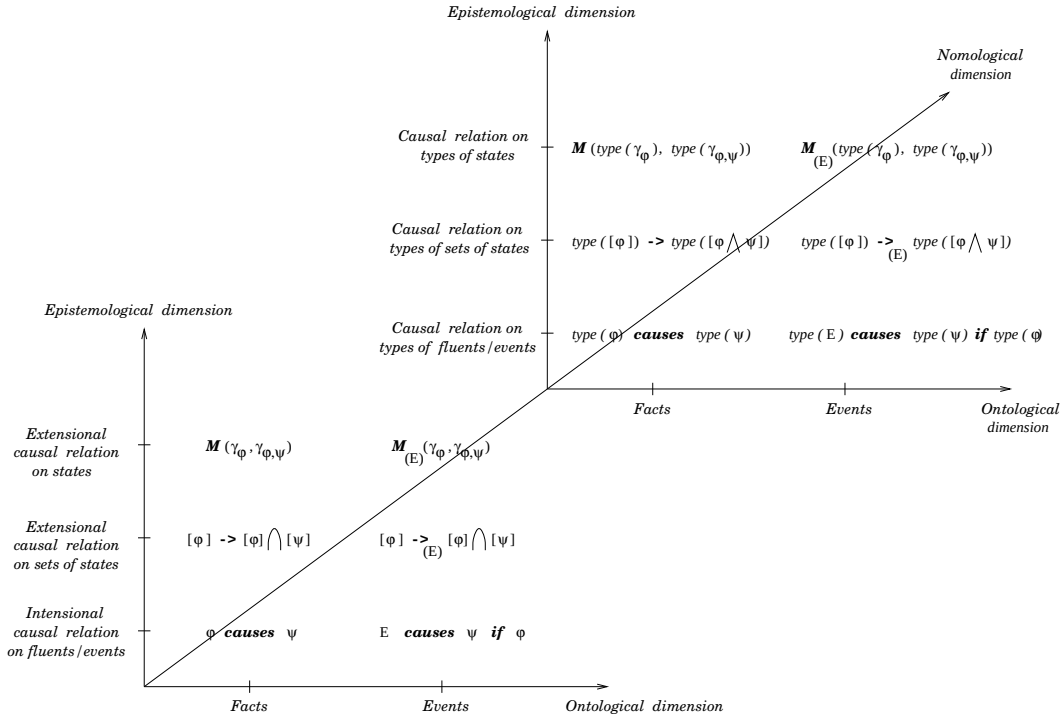

Figure 2.10: The multi-faceted causation space.

> For one might hold, first, that causal relation *presuppose* underlying causal laws — that is, that it is impossible for two states of affairs to be causally related unless they fall under some causal law — but second, that whether two states of affairs are causally related need not always be logically determined by their non-causal properties and relations, together with what causal laws there are.

In particular, this position is motivated by a possibility of nondeterministic laws ("quantum physics seems to lend strong support to the idea that the basic laws of nature may well be probabilistic" [66] and hence, indeterministic). Thus, one may entertain a view that "there can be worlds that agree with respect to, first, all of the non-causal properties of, and relations between, events, secondly, all causal laws, and, thirdly, the direction of causation, but which disagree with respect to some of the causal relations between corresponding events" [66].

To illustrate that this is not just a philosophic finesse, let us use a "simple" quantum theory example, described recently in the popular science magazine "Newton" [34]:

> Take two uranium atoms and wait. After a while one will emit an alpha particle (two protons and two neutrons). The other will not. According to our best understanding of the Universe, there is absolutely no way to predict which of the uranium atoms will decay first. We know the half-life of uranium from which we can calculate the probability of the decay but this probability is the same for both atoms. There is no activity inside the nucleus, no gears, no details, no hidden variables which, if we knew them, would allow us to predict the time of the radioactive decay. The time of the decay is unknown and by all indications unknowable. Uranium atoms decay by chance, with a certain probability, but without a cause. Radioactive decay is an event without a cause.

One may argue that, actually, it may turn out to be the case that there *are* some hidden variables involved[8]. More importantly, however, an intelligent artificial agent

[8] As mentioned in the very same article, "before radioactivity was understood, the mysterious rays emitted by uranium were called alpha particles. We now know that alpha particles are made of two protons and two neutrons."

should not be expected to exhibit uniform causal reasoning on all ontological levels. The cited "Newton" article continued with the following revealing propagation: "Alpha particles tunnel out of uranium nuclei. In nuclear power plants around the world, tunneling alpha particles heat water to steam that turns a turbine and makes 20 per cent of the world's electricity". We believe that this kind of "streamlined" reasoning about quantum, thermodynamic, economic and other various actions (that belong to different domains) sounded intuitive precisely because causation served as an approximation of some underlying laws (deterministic or otherwise), and not as a manifestation of complex and interleaving causal laws (deterministic or otherwise).

In any case, the view taken in this work is that the multi-faceted causation space (Figure 2.10) allows us to better position our semantical framework, while avoiding commitment to a particular philosophical stand-point. More precisely, the binary relation $\mathcal{M}$, defined on information states, is an epistemic construct. In addition, whenever we translate between the causal relation $\mathcal{M}$ and individual causal rules or some relation(s) defined in terms of fluent-triggered causality, we confine the translation(s) to the epistemological dimension in the ontological-epistemological plane. In other words, such translations are not intended to reflect on or discover new causal laws, but rather to re-shape causal relations used by the agent. This is one of the underlying reasons allowing us to entertain a possibility that our motivating approaches can be represented in a unifying setting: they seem to belong to the same surface in the multi-dimensional causation space.

This point is important, in our view, not only because some of the suggested terminology is misguided (for example, *causal laws* in McCain and Turner's approach do not capture type-causality and are not laws as such). More importantly, we wish to avoid confusion between causal laws (that belong to the nomological plane) and relations between epistemic states and sets of epistemic states. Consider, for example, an abstract relation $\rhd$ that links sets of world states compatible with different post-conditions of actions; in other words, $\rhd$ is defined on $2^{\mathcal{W}} \times 2^{\mathcal{W}}$. For instance, one may specify $[e_1] \rhd [e_2]$, where $[e]$ is defined as before to be a subset of states in $\mathcal{W}$ compatible with the post-conditions of action $[e]$. We do not intend to provide here an intuitive meaning for such a relation (it is not part of our semantics). We just attempt to illustrate that, although the

relation $\rhd$ captures some regularity among states (in particular, that all states compatible with $[e_1]$ are related to all states compatible with $[e_2]$), this regularity does not relate a *type* of action e_1 to a *type* of action e_2.

Von Wright offered a distinction between *sharp* and *blurred* pictures of causality [69, p. 110]. The "sharp" picture emerged when a logico-atomistic structure was proposed as a "fiction" or "model" of the world, and causal relations were defined and studied in this model. It was contrasted to the "blurred" picture of causality that is "employed in scientific practice and underlie the causal talk of natural and social scientists and historians". Not unlike von Wright, we hope that the distinctions made in the multi-faceted causation space considered in this section, may lead to a better appreciation of claims "regarding the operations of causality in the web of facts constituting reality" [69, p. 110].

2.6 A Simple Augmented Preferential Semantics

The introduction of the causal relation $\mathcal{M}$ defined on the information state-space completes our preliminary framework. The only remaining piece is a refinement of the selection function. Our motivation is to retain the preferential style of the semantics considered earlier, and make use of the causal relation in the selection function.

Intuitively, we can trace an agent's reasoning about an action e producing successor states $Res(w, e)$ as follows. First of all, some bounded start area (let us call it a *gradient*) is determined in the information state-space Γ — the agent entertains the states in the gradient area as the nearest possible information states compatible with the action's direct effects. Then, the agent begins a propagation from the gradient, driven by the causal relation $\mathcal{M}$. This propagation may explore the whole state-space Γ, but is expected to reach at least one stable state (an element of $\mathcal{K}_{\mathcal{M}}$) — otherwise the action e is qualified, and $Res(w, e) = \emptyset$. The length or configuration of the explored causal chains should not matter, as the propagation process occurs in the information state-space — in other words, a real-time aspect of knowledge dynamics is secondary to our objectives. What is important, however, is that final state(s) are "stable" and "reachable" from the gradient area. If a projection from such a final state to the state-space $\mathcal{W}$ results in a legitimate

state r compatible with e (in other words, $r \in \mathcal{D} \cap [e]$), then the state r is a successor state.

Let $\mathcal{M}^*$ be the transitive closure of the relation $\mathcal{M}$. We shall say that an information state β is $\mathcal{M}$-reachable from an information state α, if and only if $\mathcal{M}^*(\alpha, \beta)$. By $\alpha(w)$ we again denote any information state such that $\mathcal{P}(\alpha(w)) = w$. Also, let us refer to the set $min(<_{\alpha(w)}, [e]^\Gamma)$ as the gradient area — intuitively, it contains all $<_{\alpha(w)}$-minimal information states whose projections are compatible with the post-conditions of action e. Chapter 7 will elaborate on other possible gradient definitions.

We shall say that a legitimate state r is a successor state, $r \in Res(w, e)$, if and only if r is compatible with the action's direct effects and is a projection of some stable information state, which is $\mathcal{M}$-reachable from a minimal state in the gradient. More precisely,

$$Res(w, e) = \mathcal{D} \cap [e] \cap \mathcal{X}(\{\rho \in \mathcal{K}_\mathcal{M} : \ \mathcal{M}^*(\varphi, \rho), \text{ where } \varphi \in min(<_{\alpha(w)}, [e]^\Gamma)\}).$$

The set

$$\{\rho \in \mathcal{K}_\mathcal{M} : \ \mathcal{M}^*(\varphi, \rho), \text{ where } \varphi \in min(<_{\alpha(w)}, [e]^\Gamma)\}$$

used in this selection function, contains all stable states $\rho \in \mathcal{K}_\mathcal{M}$ which are $\mathcal{M}$-reachable from some minimal state φ in the gradient $min(<_{\alpha(w)}, [e]^\Gamma)$. Given our density condition that

$$\mathcal{D} \cap \mathcal{X}(\Gamma \setminus \mathcal{K}_\mathcal{M}) = \emptyset,$$

we could omit the requirement $\rho \in \mathcal{K}_\mathcal{M}$ in the definition of $Res(w, e)$ — because non-stable states would not be projected onto legitimate states in $\mathcal{D}$ anyway. But it is more intuitive to consider only stable $\mathcal{M}$-reachable states to begin with.

In any case, this definition clearly separates aspects of minimality and causality in our semantics — the former is captured by a preferential structure $\mathcal{O}$ and the latter by a binary relation $\mathcal{M}$. We claim therefore, that both minimal change and causation are essential in furnishing a concise solution to the frame and ramification problems.

In summary, our first (simple) version of an *augmented preferential semantics*, can be presented by a tuple $\langle \mathcal{W}, \ \mathcal{D}, \ \Gamma, \ \mathcal{E}, \ \mathcal{O}, \ \mathcal{M} \rangle$, where

- $\mathcal{W}$ is the set of world states;

- $\mathcal{D}$ is the set of legitimate world states;
- Γ is the set of information states;
- $\mathcal{E}$ is the set of actions;
- $\mathcal{O}$ is the preferential structure on $\Gamma \times \Gamma$ (the set of orderings $<_\alpha$ defined with respect to each information state $\alpha \in \Gamma$);
- $\mathcal{M}$ is the causal binary relation on $\Gamma \times \Gamma$,

together with the functions

- post-condition $[e] : \mathcal{E} \to 2^{\mathcal{W}}$;
- projection $\mathcal{P}(\gamma) : \Gamma \to \mathcal{W}$;
- selection $Res(w, e) : \mathcal{W} \times \mathcal{E} \to 2^{\mathcal{W}}$.

Let us briefly analyse, at this stage, an extreme possibility that $\mathcal{D} = \mathcal{W}$, meaning that, whatever underlying (non-causal) laws apply in the domain under consideration, they do not constrain the set of world states $\mathcal{W}$ at all (more intuitively, this particular domain is anomic)[9]. In this case, the density condition

$$\mathcal{D} \cap \mathcal{X}(\Gamma \setminus \mathcal{K}_{\mathcal{M}}) = \emptyset,$$

ensures that

$$\mathcal{W} \cap \mathcal{X}(\Gamma \setminus \mathcal{K}_{\mathcal{M}}) = \emptyset.$$

Furthermore, using properties of set-projection $\mathcal{X}$, we recall that

$$\mathcal{W} = \mathcal{X}(\mathcal{K}_{\mathcal{M}}) \cup \mathcal{X}(\Gamma \setminus \mathcal{K}_{\mathcal{M}}),$$

leading to the conclusion

$$\mathcal{W} = \mathcal{X}(\mathcal{K}_{\mathcal{M}}).$$

In other words, every world state is a projection of some stable information state, and not an unstable state.

[9] Although we specified earlier, in Section 2.2.1, that $\mathcal{D} \subset \mathcal{W}$, the curious possibility $\mathcal{D} = \mathcal{W}$ deserves some attention.

Therefore, there cannot be any unstable information states — otherwise, such a state would have to be projected "outside" of the world state-space. Hence,

$$\Gamma = \mathcal{K}_{\mathcal{M}},$$

meaning that every (stable) state is *trivially* stable — it is not related by means of the relation $\mathcal{M}$ with any other information state (neither as cause nor as effect). We can conclude that in this anomic domain there cannot be any causal relations at all, and the relation $\mathcal{M}$ is empty.

Another distinct class of domains contains domains where all legitimate states are projections of only trivially stable information states, although the relation $\mathcal{M}$ does not have to be empty. In other words, stable states β that appear as effects in pairs $\mathcal{M}(\alpha, \beta)$ are always projected in this domain onto illegitimate states. We shall refer to such domains as *causally trivial* domains — because there is no way to use causality in manipulating legitimate states. On the contrary, in a *causally non-trivial domain*, there is at least one legitimate world state that is a projection of a non trivially stable state.

One more interesting class of domains can be specified by setting

$$\mathcal{D} = \mathcal{X}(\mathcal{K}_{\mathcal{M}}).$$

This extreme case of the density condition is referred to as the *compactness* property and is discussed further in Chapter 7.

Further chapters will also highlight another interesting aspect of propagation-driven ramifications: *context-sensitivity*, resulting in a multiplicity of possible gradients, and a means for their choice. However, we believe that this aspect can only be analysed after a thorough analysis of our motivating approaches. The current chapter merely sets a framework, and hence, we shall follow only with brief outlines of these approaches and a sketch of a *general augmented preferential semantics*.

2.7 McCain and Turner's Causal Fixed-points Approach

In this section we sketch McCain and Turner's [37] causal theory of actions. In so doing we shall introduce some further notation that will be useful for the remainder of the

work. Let $\mathcal{F}$ be a finite set of symbols from a fixed language $\mathcal{B}$, called fluent names. A *fluent literal* is either a fluent name $f \in \mathcal{F}$ or its negation, denoted by $\neg f$. Let $L_{\mathcal{F}}$ be the set of all fluent literals defined over the set of fluent names $\mathcal{F}$. A maximal consistent set of fluent literals is called a *state*. For convenience, we will denote the set of all states as $\mathcal{W}$ (identifying it with the set of world states in our framework). We shall call the number m of fluent names in $\mathcal{F}$ the dimension of $\mathcal{W}$. By $[\phi]$ we denote all states consistent with the sentence $\phi \in \mathcal{B}$ (i.e., $[\phi] = \{w \in \mathcal{W} : w \vdash \phi\}$). Obviously, our intention is to identify $[\phi]$ with the post-conditions of the action specifying ϕ as its direct effects. Domain constraints are sentences which have to be satisfied in all states.

McCain and Turner introduce a new connective $\Rightarrow$ to denote a causal relationship between sentences ϕ and ψ of the underlying language $\mathcal{B}$. This allows for expressions of the form $\phi \Rightarrow \psi$ (where $\phi,\ \psi \in \mathcal{B}$) which are termed *causal laws* (or *causal rules* — we prefer the latter term, for the reasons mentioned earlier). Nesting of $\Rightarrow$ is not permitted. For the sake of simplicity we shall assume here that the antecedent of any causal rule is consistent. A set of causal rules $\mathcal{Q}$ is referred to as a *causal system*. Given any set of sentences $\Lambda \subseteq \mathcal{B}$ and a causal system $\mathcal{Q}$, the (causal) *closure* of Λ in $\mathcal{Q}$ is denoted $C_{\mathcal{Q}}(\Lambda)$ and defined to be the smallest superset of Λ closed under classical logical consequence such that for any $\phi \Rightarrow \psi \in \mathcal{Q}$, if $\phi \in C_{\mathcal{Q}}(\Lambda)$, then $\psi \in C_{\mathcal{Q}}(\Lambda)$. We also say that Λ *causally implies* ϕ with respect to $\mathcal{Q}$ if and only if $\phi \in C_{\mathcal{Q}}(\Lambda)$ and denote this as $\Lambda \mid\!\sim_{\mathcal{Q}} \phi$. Any state r is legitimate with respect to $\mathcal{Q}$ if and only if $r = C_{\mathcal{Q}}(r) \cap L_{\mathcal{F}}$. That is, a state is legitimate if and only if it does not contravene any causal laws of $\mathcal{Q}$.

McCain and Turner's aim is to determine the set of possible resultant (or successor) states $Res_{\mathcal{Q}}(w, e)$ given an initial state w and the direct effects (or post-conditions) of an action represented by the sentence e.[10] Formally speaking, we have for any causal system $\mathcal{Q}$ a function $Res_{\mathcal{Q}}$ mapping a legitimate (initial) state w and sentence e (direct effects) to the set of states $Res_{\mathcal{Q}}(w, e)$ according to the definition [37]:

$$r \in Res_{\mathcal{Q}}(w, e) \text{ iff } r = \{p \in L_{\mathcal{F}} :\ (w \cap r) \cup \{e\} \mid\!\sim_{\mathcal{Q}} p\}$$

We often refer to the elements of $Res_{\mathcal{Q}}(w, e)$ as *causal fixed-points*. Note that it follows

[10] Here we refer to actions only through their direct effects as actions play no direct role in McCain and Turner's framework.

from this definition that if $r \in Res_{\mathcal{Q}}(w, e)$, then $r \in [e]$ (i.e., r must satisfy the direct effects of the action). Intuitively speaking, the elements of $Res_{\mathcal{Q}}(w, e)$ are simply those e-states where all changes with respect to w can be justified by the underlying causal system. In short, every literal in the successor state must be "explained" either as persisting through the action, or as a direct effect, or as a causal ramification of other literals of the successor state.

It was emphasised previously that this definition captures the causal minimisation policy: the world changes as little as *necessary* when an action is performed. Clearly, we would need to clarify exactly where and how the Principles of Minimal and Causal Change interact in the definition of $Res_{\mathcal{Q}}(w, e)$, before providing a preferential-style semantics (augmented or otherwise) for causal fixed-points. This will be done in Chapter 4. Here, however, we just make a very promising observation:

$$Res_0(w, e) \subseteq Res_{\mathcal{Q}}(w, e) \subseteq Res_1(w, e),$$

for every state $w \in \mathcal{W}$ and action e, if the selection functions $Res_0(w, e)$ and $Res_1(w, e)$ employ the PMA ordering. In other words, a causal fixed-point always lies "between" the first and second boundaries of successor states. Figure 2.11 revisits the earlier example of Figure 2.7, and illustrates potential causal fixed-points in the area bounded by a dotted line.

2.8 Thielscher's Approach of Causal Relationships

In this section we briefly review Thielscher's [63] causal relationships approach, while reusing some notation from the previous section. We will adopt from Thielscher [63] the following notation. If $\epsilon \in L_{\mathcal{F}}$, then $|\epsilon|$ denotes its *affirmative component*, that is, $|f| = |\neg f| = f$, where $f \in \mathcal{F}$. This notation can be extended to sets of fluent literals as follows: $|S| = \{|f| : f \in S\}$.

Thielscher's [63] causal theory of action consists of two main components: *action laws* which describe the direct effects of actions performed in a given state, and *causal relationships* which determine the indirect effects of actions. Every action law[11] contains

[11] Again, action laws should be more appropriately called action rules.

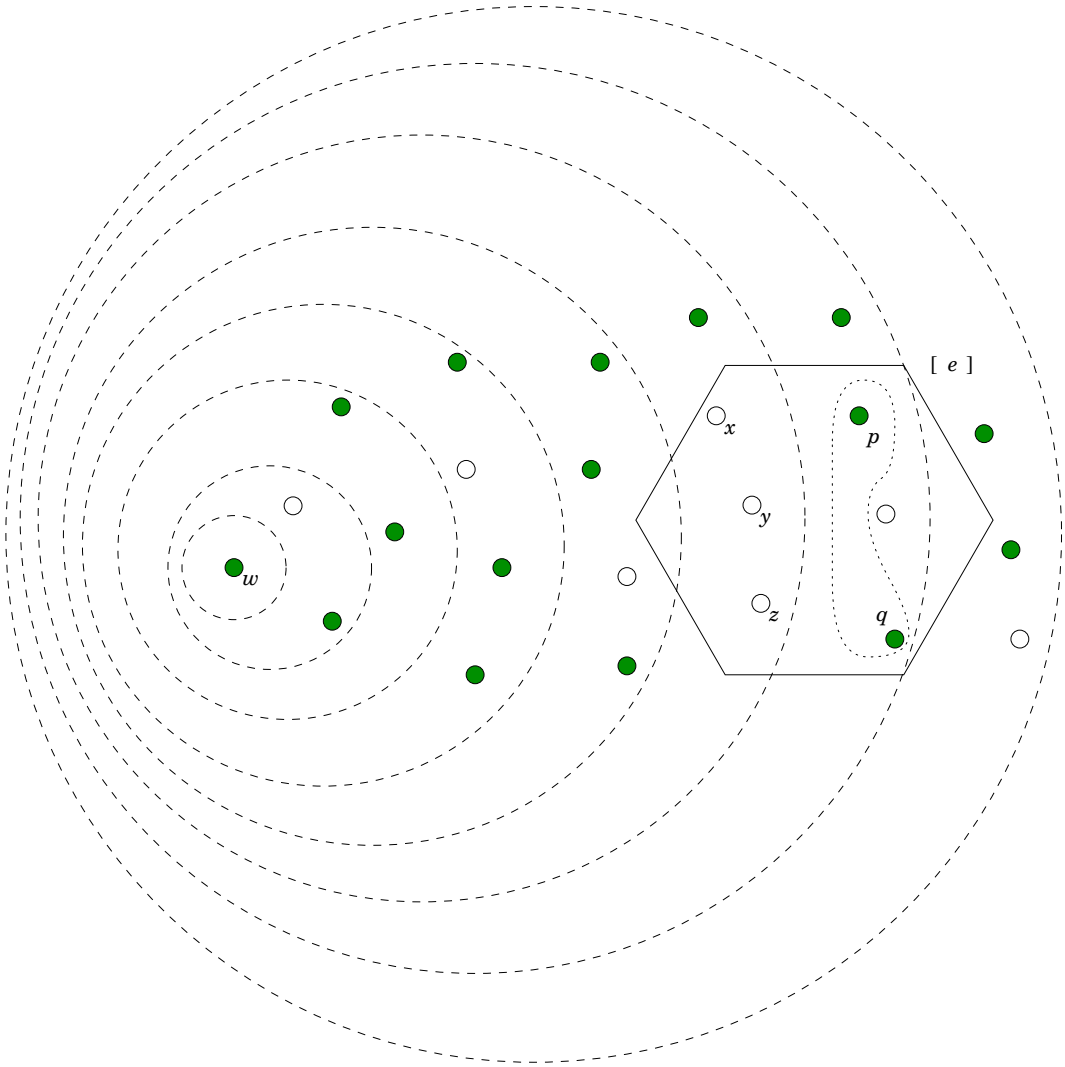

Figure 2.11: Only the states p and q, lying on the second boundary, can be causal fixed-points: $Res_{\mathcal{Q}}(w, e) \subseteq \{p, q\}$.

a condition C, which is a set of fluent literals, all of which must be contained in an initial state where the action is intended to be applied; and a (direct) effect E, which is also a set of fluent literals, all of which must hold in the resulting state after having applied the action. Condition and effect are constructed from the same set of fluent names so that the state obtained from a direct effect is determined by removing C from the initial state and adding E to the result. An action may result in a number of state transitions.

Definition 2.8.1 *Let $\mathcal{F}$ be the set of fluent names and let $\mathcal{A}$ be a finite set of symbols called action names, such that $\mathcal{F} \cap \mathcal{A} = \emptyset$. An action law is a triple $\langle C, a, E\rangle$ where C, called* condition*, and E, called* effect*, are individually consistent sets of fluent literals, composed of the very same set of fluent names (i.e., $|C| = |E|$) and $a \in \mathcal{A}$. If w is a state, then an action law $\alpha = \langle C, a, E\rangle$ is applicable in w iff $C \subseteq w$. The application of α to w yields the state $(w \setminus C) \cup E$ (where $\setminus$ denotes set subtraction).*

Causal relationships are specified as ϵ *causes* ρ *if* Φ, where ϵ and ρ are fluent literals and Φ is a fluent formula based on the set of fluent names $\mathcal{F}$. Another important instrumental concept is a state-effect pair (s, E) containing a state s and a set of fluent literals E. The second component is used to "record" a (partial) history of fluents that changed their values in transitions leading to the state represented by the first component of the pair — more precisely, the "history" contains only the latest (current) values of all the changed fluents, making it a snapshot of current effects.

Definition 2.8.2 *Let (s, E) be a pair consisting of a state s and a set of fluent literals E. Then a causal relationship ϵ* causes *ρ if Φ is applicable to (s, E) iff $\Phi \wedge \neg\rho$ is true in s, and $\epsilon \in E$. Its application yields the pair (s', E'), denoted as $(s, E) \rightsquigarrow (s', E')$, where $s' = (s \setminus \{\neg\rho\}) \cup \{\rho\}$ and $E' = (E \setminus \{\neg\rho\}) \cup \{\rho\}$.*

In other words, a causal relationship is applicable if Φ holds at the current state s, the indirect effect ρ is false and the cause ϵ is among the current effects E. Note that ϵ must be among the current effects; being true at the current state is not sufficient.

A possible *successor state* is determined through the repeated application of causal relationships. In so doing we may temporarily end up in states violating domain constraints. This is permissible only if subsequent applications of causal laws result in legal

states. Specifically, given an initial state w and action a, the set of possible successor states $Res_{RD\mathcal{L}}(w, a)$ is determined as follows.

Definition 2.8.3 *Let $\mathcal{F}$ be the set of fluent names, A a set of action names, $\mathcal{L}$ a set of action laws, D a set of domain constraints, and R a set of causal relationships. Furthermore, let w be a state satisfying D and let $a \in A$ be an action name. A state r is a* successor state *of w and a, denoted $r \in Res_{RD\mathcal{L}}(w, a)$, iff there exists an applicable (with respect to w) action law $\alpha = \langle C, a, E\rangle \in \mathcal{L}$ such that*

1. $((w \setminus C) \cup E, E) \overset{*}{\rightsquigarrow} (r, E')$ *for some E', and*
2. *r satisfies D,*

where $\overset{*}{\rightsquigarrow}$ denotes the transitive closure of $\rightsquigarrow$.

As mentioned before, the occurrence of a literal ϵ in a state s does not guarantee that a causal relationship ϵ *causes* ρ *if* Φ is applicable to a pair (s, E) — to ensure applicability, the literal ϵ has to belong to the current effects E. That is why, in order to trace causal propagation with causal relationships, one needs to keep an explicit (and changing) account of context-dependent effects of actions.

The description in this section is only an outline, and the causal relationship approach will be further investigated in Chapter 5. At this stage we just observe (following Thielscher [63]) that, given a simple construction of causal relationships from causal rules, every causal fixed-point is a successor according to the causal relationship approach (but not vice versa).

2.9 Sandewall's Causal Propagation Semantics

The causal propagation semantics[12] introduced by Sandewall [56], uses the following basic concepts (we deliberately change some notation in order to provide better links with our framework). The set of possible states of the world, formed as the Power-set of state variables, is denoted as $\mathcal{W}$. $\mathcal{E}$ is the set of possible actions. The causal propagation

[12] In some earlier papers it was called the *transition cascade semantics*.

semantics extends a basic state transition semantics with a *causal transition relation*. The causal transition relation C is a non-reflexive relation on states in $\mathcal{W}$. A state r is called *stable* if it does not have any successor s such that $C(r, s)$; we will denote the set of stable world states $\{r \in \mathcal{W} : \neg\exists s \in \mathcal{W}, C(r, s)\}$ as $\mathcal{S}_c$. Another component, $\mathcal{D}$, is a set of admitted states chosen as a subset of $\mathcal{S}_c$. The set $\mathcal{D}$ may or may not be chosen to contain all the stable states — in other words, some stable states may not be admitted.

Another important concept, introduced by Sandewall, is an *action invocation* relation $G(e, r, r')$, where $e \in \mathcal{E}$ is an action, r is the state where the action e is invoked, and r' is "the new state where the instrumental part of the action has been executed" [56]. In other words, the state r' satisfies the direct effects of the action e. It is required that every action is always invokable, that is, for every $e \in \mathcal{E}$ and $r \in \mathcal{W}$ there must be at least one r' such that $G(e, r, r')$ holds. Of course, this requirement does not mean to guarantee that every action results in an admitted state — on the contrary, the intention is to trace the indirect effects of the action, possibly reaching an admitted (and, therefore, stable) state.

A finite (the infinite case is omitted) transition chain for a state $w \in \mathcal{D}$ and an action $e \in \mathcal{E}$ is a finite sequence of states $r_1, r_2, ..., (r_k)$, where $G(e, w, r_1)$ and $C(r_i, r_{i+1})$ for every $i, 1 \leq i < k$, and where r_k is a stable state. The last element of a finite transition chain is called a result state of action e performed in state w.

These basic concepts define an *action system* as a tuple $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G \rangle$. The following definition strengthens action systems based on the causal propagation semantics.

Definition 2.9.1 *If three states w, p, q are given, we say that the pair p, q respects w, denoted as $\lhd_w(p, q)$, if and only if $p(f) \neq q(f)$ implies $p(f) = w(f)$ for every state variable f that is defined in $\mathcal{W}$, where $r(f)$ is the valuation of variable f in state r.*

An action system $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G \rangle$ is called respectful if and only if, for every $w \in \mathcal{D}$, every $e \in \mathcal{E}$, w is respected by every pair r_i, r_{i+1} in every transition chain (for the state w), and the last element of the chain is a member of $\mathcal{D}$.

According to Sandewall [56], respectful action systems are intended to ensure that in each transition there cannot be changes in state variables which have changed previously upon invocation or in the causal propagation sequence. This requirement, of course,

guarantees that a resultant state is always consistent with the direct effects of the action (which cannot be cancelled by indirect ones), and that there are no cycles in transition chains.

As with many other state transition action systems, the intention is to characterise a result state in terms of an initial state w and action e, without "referring explicitly to the details of the intermediate states" [56]. In other words, it is desirable to define a selection function $Res(w, a)$. For a respectful action system $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G \rangle$, a selection function can be given as

$$Res_{C\mathcal{D}G}(w, e) = \{r_k \in \mathcal{D} : G(w, e, r_1), C(r_i, r_{i+1}), \lhd_w(r_i, r_{i+1}), 1 \leq i < k\}.$$

While there are certain similarities between the causal propagation semantics and our augmented preferential semantics (including components such as $\mathcal{W}$, $\mathcal{E}$, $\mathcal{D}$, and the causal transition relation), it is not obvious that they define the same successor states. In particular, the Principal of Minimal Change is not explicit in the causal propagation semantics. In Chapter 6 we shall discover what role is played by minimality in this semantics, being non-trivially masked by the invocation relation.

2.10 Towards a General Augmented Preferential Semantics

Conciseness and intuitive representation were two primary objectives motivating our simple augmented preferential semantics outlined earlier. It is possible to demonstrate that such a semantics can be generalised and cover the motivating approaches, while staying on a technical logical level. However, it would be beneficial to show that there are common assumptions made in these different proposals on a philosophic level — in particular, with respect to causation. Then, a unified semantics would provide additional insights into the views on causation, shared by these approaches.

2.10.1 Direction of Causation and Causal Priority

One particular property of causation that is definitely accepted by the three approaches under consideration (and by a clear majority of other accounts), is *unidirectionality* of

causation. This property is reflected in the asymmetry condition ($\mathcal{M}_2$): if the (information) state of affairs α causes the (information) state of affairs β, then it cannot be the case that β causes α.

Another important aspect of causal relations is that they define an ordering of *causal priority* [65]. In other words, causes are prior to their effects and, if one presupposes a temporal ordering, causes cannot succeed their effects in time.

There are, of course, some accounts that disagree that causes must be prior to their effects. For example, Lewis [25] argued against this property, because "it rejects *a priori* certain legitimate physical hypotheses that posit backwards or simultaneous causation". This position was criticised by Shoham [60] on the grounds that "our intuitions about causation seem intimately bound to temporal precedence", and "we should be wary of philosophical theories which flatly contradict human intuition, especially when dealing with concepts which we use regularly in everyday life". Let us elaborate on this point further, with the remarks, made in a recent Umberto Eco essay [10, p. 184], on "the sorts of reasoning we find in Phaedrus's fable of the wolf and the lamb, in which it is said that the lamb, by drinking downstream of the wolf, disturbs him as he is drinking upstream". Eco argues that

> I am not claiming that there is only one arrow of time. There may be more than we believe. Thinking along the same lines, I would not claim that there is one sort of geometry, Euclidian geometry, because there are many other sorts. All I would claim is that in daily life, when we have to hang a picture on a wall, we must follow Euclidian geometry and not Lobachevsky's, and if we ask at what time the fast train from Paris which left at seven o'clock will arrive in Lyon, we have to think in terms of the time which our clocks obey and not in terms of Bergson's internal time consciousness

and follows with a pragmatic conclusion: "This leads me to pronounce judgement: the lamb cannot disturb the wolf's water. Case closed."

Causal priority is not necessarily captured by causal asymmetry as, for example, pointed out by Tooley [65, p. 179]. Tooley considers the ancestral of causal relation that is necessarily transitive (in addition to being asymmetric), and a strict partial ordering

defined by this relation on states of affairs:

> But if R is any asymmetric and transitive relation, then the inverse relation, R^*, is also asymmetric and transitive, and it generates an ordering that is indistinguishable from that generated by R, except that it is opposite in direction. This means that any satisfactory account of causal priority, in addition to explaining the asymmetry of causal relations, must also supply some account of why it is the *direction* defined by those relations, rather than that defined by the inverse relations, that is *the* direction of causal processes.

Our motivating approaches do not explicitly use a temporal ordering, and therefore, we would like to capture the property of causal priority without relying on some notion of temporal priority. Moreover, many philosophers have argued that causal concepts are more basic than temporal ones, and have tried to analyse the latter in terms of the former. For instance, D. H. Mellor [42] posed the question: "The cause-effect relation has no preferred spatial dimension. Why then does it have a temporal one?", and argued that "causation is what distinguishes time from space and gives it its direction: in short, that time is the causal dimension of spacetime".

Fortunately, our framework already includes a condition that may help us in identifying the causal priority (at least, for non causally trivial domains). This surprisingly helpful condition is the requirement that

$$\mathcal{D} \cap \mathcal{X}(\Gamma \setminus \mathcal{K}_{\mathcal{M}}) = \emptyset,$$

Let us show that this density condition affects causal priority — in particular, it prevents the reversal described by Tooley [65]. Consider a causally non-trivial domain[13]. Since the domain is not causally trivial, there is at least one legitimate world state that is a projection of a non trivially stable state. Let $w \in \mathcal{D}$ be such a state. In other words, there must be at least two information states α and β, such that $\mathcal{M}(\alpha, \beta)$, where $\beta \in \mathcal{K}_{\mathcal{M}}$, and $\mathcal{P}(\beta) = w$.

[13]To reiterate, a non trivially stable state is related by means of the relation $\mathcal{M}$ with some other information state; in a causally non-trivial domain there is at least one legitimate world state that is a projection of a non trivially stable state.

Let the relation $\mathcal{M}^{\times}$ be the inverse of $\mathcal{M}$. Hence $\mathcal{M}^{\times}(\beta, \alpha)$, and β is not a stable state with respect to $\mathcal{M}^{\times}$. We obtain that $\beta \notin \mathcal{K}_{\mathcal{M}^{\times}}$, i.e. $\beta \in \Gamma \setminus \mathcal{K}_{\mathcal{M}^{\times}}$. If causal priority is not enforced (i.e., the inverse relation is as causal as the original), then the relation $\mathcal{M}^{\times}$ should also satisfy the requirement

$$\mathcal{D} \cap \mathcal{X}(\Gamma \setminus \mathcal{K}_{\mathcal{M}^{\times}}) = \emptyset.$$

This condition and the observation that $\beta \in \Gamma \setminus \mathcal{K}_{\mathcal{M}^{\times}}$ lead to the conclusion that $\mathcal{P}(\beta) \notin \mathcal{D}$, or in other words, $w = \mathcal{P}(\beta)$ is not a legitimate state. This is a contradiction with our assumption that $w \in \mathcal{D}$. In other words, the density condition

$$\mathcal{D} \cap \mathcal{X}(\Gamma \setminus \mathcal{K}_{\mathcal{M}^{\times}}) = \emptyset$$

where $\mathcal{M}^{\times}$ is the inverse relation for a given $\mathcal{M}$, can not hold in a non causally trivial domain.

This fact distinguishes relations $\mathcal{M}$ and $\mathcal{M}^{\times}$, and establishes causal priority. Put simply, causal priority would not matter in causally trivial domains, because then reversals of causal relations would not affect legitimate states and selection functions.

2.10.2 Causal Propagation and Context-sensitivity of Causation

A closely related topic is the notion of *causal efficacy*: not only are causes prior to their effects but they also "produce" their effects. Following von Wright [69, p. 118]:

> What makes p a cause-factor relative to the effect-factor q is, I shall maintain, the fact that by *manipulating* p, i.e., by producing changes in it 'at will' as we say, we could bring about changes in q. This applies both to cause-factors which are sufficient and those which are necessary conditions of the corresponding effect-factor.

Causal efficacy and other properties of causation, such as irreflexivity, unidirectionality (asymmetry), and priority, provide grounds for causal propagation along the relation $\mathcal{M}$. In other words, the intuition behind $\mathcal{M}(\alpha, \beta)$ and $\mathcal{M}(\beta, \gamma)$ is that by triggering the information state α, we could bring about changes reflected in the information state β, and that, in turn, produces changes represented in γ, and so on.

One particular shortcoming of causal unidirectionality is that it "disregards the fact that all known actions are accompanied or followed by reactions, that is, that the effect always reacts back on the input unless the latter has ceased to exist" [5]. As mentioned by Bunge in the same work, causality may be a good approximation when there is a strong dependence of the effect upon its cause, with a negligible backward reaction, and "whenever the cause has ceased existing". However, the point that we pursue is that causal propagation in the (higher-dimensional) information state-space is precisely the tool enabling better approximation, and so it could be completely free of backward reaction. In short, in terms of the relation $\mathcal{M}$, an information state should not react back upon a previous transition state (negligibly or otherwise) — if there is a reaction to be accounted for, it should be represented by other acyclic causal chains in the information state-space.

An appealing analysis of necessity and sufficiency of causes with respect to their effects was given by Mackie [36]. Although this account is often criticised, it certainly captures the problem intuitively. Instead of defining causes as necessary and/or sufficient conditions, Mackie proposes that, in the relation 'c causes e', the cause c can be typically represented as "an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result" [36] — the INUS condition. More precisely, 'c causes e' whenever $(c \wedge x) \vee y$ is a necessary and sufficient condition for e, for some x and y, where neither c, x or y is redundant. A frequently cited example, suggested by Mackie in the original work, is the example of a short circuit said to have caused a fire, in the presence of inflammable material, the absence of suitably placed sprinkler, and other relevant conditions. These unnecessary conditions x are sufficient when combined with the short circuit c, which by itself is insufficient. The fire can start in other ways y as well (eg., "overturning of a lighted oil stove") — that is why $(c \wedge x) \vee y$ becomes necessary and sufficient.

Shoham [60] supported the intuitive distinction between causes c and other relevant enabling conditions x, but criticised the INUS condition as being too weak formally. The main difference between Mackie's account and Shoham's well-known formalisation, *the logic of chronological ignorance*, was described by Shoham as follows:

> ... through the use of modal logic, I made those other conditions secondary: whereas causes need to be known (i.e. they are $\Box$-conditions), it is sufficient that the "enabling" conditions not be known to be false (i.e., they are $\Diamond$-conditions).

At this stage, we would like to point out that causal rules $\Box c \wedge \Diamond x \supset \Box e$, employed by the logic of chronological ignorance, or INUS conditions in general, presuppose causal propagation — in some sense. Mackie mentioned "some line or chain of causation, some continuous causal process" [35], while the logic of chronological ignorance used "an effective procedure ... starting with knowledge of only tautologies prior to the earliest point ... and then iteratively progressing in time, adding only knowledge ... that is necessitated by prior knowledge and ignorance", while skipping irrelevant points [59].

We maintain that causal propagation is employed, as a process exploring the information state-space in a search for necessary and/or sufficient conditions, in all our motivating approaches, outlined above in sections 2.7 — 2.9. This might seem to be more obvious in Sandewall's and Thielscher's approaches, and far less so in McCain and Turner's formalisation. Nevertheless, we intend to show that forces producing successor states are quite similar in all these frameworks. While the similarities can be explained by common assumptions with respect to the nature of causality and minimality, the differences, we believe, are due to different ways to handle the *context-sensitive* nature of causal propagation.

Causal propagation is a powerful mechanism, enabling efficient *uniform* reasoning about actions and their ramifications. However, a propagation process typically takes place against a background, a *causal field*. The notion of causal field is well-known — it was introduced by Anderson [1], and used by Mackie [36]. In Reasoning about Action, causal field is usually understood as "the causal theory, relative to which causal relations either do or do not hold" [60].

It would be surprising if the causal theories maintained by the three approaches under consideration set identical "causal fields". While all these approaches treat causation as a context-sensitive (*indexical*) phenomenon, they use causal context differently.

The causal theories of McCain and Turner "warp" simple propagation by restricting successor states to causal fixed-points. The causal relationships of Thielscher create causal histories that affect simple propagation, because the process must keep an account of all changes. The causal propagation of Sandewall is limited to respectful systems (trajectories) in a sensitive way as well.

We propose to capture context-sensitivity by specifying and choosing different gradient functions, employed by our semantics. These functions would combine and blend causality and minimality in various ways. Some gradient choices would capture contextual information leading exclusively to fixed-points, some would allow an agent to propagate in the information state-space that embeds causal histories, and some would constrain minimality metrics leaving only respectful trajectories possible.

We shall describe these and other gradient choices later — in Chapter 7, where a general unifying semantics based on preferential and causal relations (defined on the information state-space) will be proposed.

2.10.3 Summary of Primitives and Notation

In this section, we summarise some of the terminology and notation that will prove useful throughout this dissertation.

Syntax-dependent definitions

- $\mathcal{F}$ is a finite set of symbols from a fixed language $\mathcal{B}$, called fluent names.
- A fluent literal is either a fluent name $f \in \mathcal{F}$ or its negation, denoted by $\neg f$.
- $L_{\mathcal{F}}$ is the set of all fluent literals defined over the set of fluent names $\mathcal{F}$.
- A set s of literals is consistent if and only if it does not contain both f and $\neg f$, for some literal f.
- A state is a maximal consistent set of fluent literals.
- The set of all states is denoted $\mathcal{W}$.
- The number of fluent names in $\mathcal{F}$ is the dimension of the set of all states $\mathcal{W}$.

- By $[\phi]$ we denote all states consistent with the sentence $\phi \in \mathcal{B}$:

$$[\phi] = \{w \in \mathcal{W} : w \vdash \phi\}.$$

- Domain constraints are sentences which have to be satisfied in all states.
- If $\epsilon \in L_{\mathcal{F}}$, then $|\epsilon|$ denotes its affirmative component, that is, $|f| = |\neg f| = f$, where $f \in \mathcal{F}$.
- For a set s of fluent literals, $|s| = \{|f| : f \in s\}$.
- A connective $\Rightarrow$ denotes a causal rule between sentences ϕ and ψ of the underlying language $\mathcal{B}$.
- A set of causal rules $\mathcal{Q}$ is a causal system.
- For a set of sentences $\Lambda \subseteq \mathcal{B}$ and a causal system $\mathcal{Q}$, the causal closure of Λ in $\mathcal{Q}$, denoted $C_{\mathcal{Q}}(\Lambda)$, is the smallest superset of Λ closed under classical logical consequence such that for any $\phi \Rightarrow \psi \in \mathcal{Q}$, if $\phi \in C_{\mathcal{Q}}(\Lambda)$, then $\psi \in C_{\mathcal{Q}}(\Lambda)$.
- Λ causally implies ϕ with respect to $\mathcal{Q}$ (denoted as $\Lambda \mid\!\sim_{\mathcal{Q}} \phi$), if and only if $\phi \in C_{\mathcal{Q}}(\Lambda)$.
- A causal relationship is specified as ϵ *causes* ρ *if* Φ, where ϵ and ρ are fluent literals and Φ is a fluent formula based on the set of fluent names $\mathcal{F}$.
- A causal relationship ϵ *causes* ρ *if* Φ is applicable to (s, E) if and only if $\Phi \wedge \neg\rho$ is true in s, and $\epsilon \in E$. Its application yields the pair (s', E'), denoted as

$$(s, E) \rightsquigarrow (s', E'),$$

  where $s' = (s \setminus \{\neg\rho\}) \cup \{\rho\}$ and $E' = (E \setminus \{\neg\rho\}) \cup \{\rho\}$.
- $\overset{*}{\rightsquigarrow}$ denotes the transitive closure of $\rightsquigarrow$.
- For states w, p, q, the pair p, q respects w, denoted as $\lhd_w(p, q)$, if and only if

$$p(f) \neq q(f) \text{ implies } p(f) = w(f)$$

for every fluent f, where $r(f)$ is a valuation of fluent f in state r.

- The symmetric difference between two states x and y is the set $\textit{Diff}(x, y) = (x \setminus y) \cup (y \setminus x)$.
- A state y is preferred to a state z in terms of the PMA ordering $\prec_x$, denoted $y \prec_x z$, if and only if $\textit{Diff}(x, y) \subseteq \textit{Diff}(x, z)$.

Syntax-independent definitions

- $\mathcal{W}$ is the finite set of world states.
- $\mathcal{D}$ is the finite set of legitimate world states.
- Γ is the finite set of information states.
- $\mathcal{E}$ is the finite set of actions.
- $\mathcal{O}$ is the preferential structure on $\Gamma \times \Gamma$ (the set of orderings $<_\alpha$ defined with respect to each information state $\alpha \in \Gamma$).
- $\mathcal{M}$ is the causal binary relation on $\Gamma \times \Gamma$.
- $\mathcal{M}^*$ is the transitive closure of the relation $\mathcal{M}$.
- $\mathcal{K}_\mathcal{M}$ is the finite set of stable information states

$$\{p \in \Gamma : \neg\exists q \in \Gamma, \mathcal{M}(p, q)\}.$$

- Post-condition function $[e] : \mathcal{E} \to 2^\mathcal{W}$.
- Selection function $Res(w, e) : \mathcal{W} \times \mathcal{E} \to 2^\mathcal{W}$.
- Projection function $\mathcal{P}(\gamma) : \Gamma \to \mathcal{W}$.
- Set-projection function $\mathcal{X} : 2^\Gamma \to 2^\mathcal{W}$ is defined as follows:

$$\mathcal{X}(\{\gamma_1, ..., \gamma_n\}) = \{\mathcal{P}(\gamma_1)\} \cup ... \cup \{\mathcal{P}(\gamma_n)\}.$$

- $\mathcal{D}^{\Gamma}$ is defined as $\{\gamma \in \Gamma : \mathcal{P}(\gamma) \in \mathcal{D}\}$.
- $[e]^{\Gamma}$ is defined as $\{\gamma \in \Gamma : \mathcal{P}(\gamma) \in [e]\}$.
- For a set A, we define $min(<_r, A)$ as a subset of A containing states nearest to the state r in terms of the ordering $<_r$:

$$min(<_r, A) = \{p \in A : \neg\exists q \in A,\ q \neq p,\ q <_r p\}.$$

- A state s is $<_r$-minimal in A if and only if $s \in min(<_r, A)$.

Chapter 3

Inertia and Causality in Action Languages

Action theories recently developed in the framework of action languages with inertia and ramifications [15, 19, 18, 21, 23, 31, 67] not only adopt the Principle of Minimal Change reinforced with the policy of categorisation (assigning different degrees of inertia to language elements), but also try to incorporate background causal knowledge. In this chapter we aim to trace the evolution of action languages and to explore interactions between diverse characteristics of action domains such as inertia and causality. This analysis should clarify how possible solutions to the Frame and the Ramification problems are affected by applying the policy of categorisation to causal domains. We first attempt to identify conditions which preserve the meaning of domain descriptions when moving across various analysed languages. Relaxing these restrictions can help in evaluating the role of the frame concept (and policy of categorisation, in general) in action languages with causality. This investigation, as mentioned earlier, is not related directly to our search of a unifying semantics, but is a necessary step permitting us to dismiss a particular categorisation policy in action theories operating with causality.

3.1 Background

As was discussed earlier, an adequate reasoning system should be capable of inferring both direct and indirect consequences of an action, as well as preserve inertial properties of the domain in question — in other words, it should resolve the Frame and Ramifica-

tion problems. Various logic-based investigations of these problems have demonstrated that in order to deduce the (indirect) effects of actions it is necessary to *efficiently* represent background domain knowledge. For instance, in the framework of first-order logic and its non-monotonic extensions, representation of background knowledge in the form of domain constraints [17] is more economical than explicit description of actions consequences such as in STRIPS [11].

Action languages with inertia and ramifications [23, 18, 22, 19] explicitly use the idea of minimising change in order to deduce the set of possible next states (successor states). The notion of minimal change is usually defined by set inclusion and incorporates the concept of frame, assigning different degrees of inertia to language elements (fluents, literals, formulae, etc.). For example, in [18] it is noted that "if F is not a frame fluent then it is not expected to keep its old value after performing an action, so that the change in its value is disregarded". In addition, as we have observed in Chapter 2, some action theories embody background information in the form of domain "causal laws" — it was successfully argued in many publications [4, 33, 37, 67] that, in general, propositions embracing causal dependencies are "more expressive than traditional state constraints" [67].

Following the discussion in the previous chapter on action-triggered and fluent-triggered causality, we shall try, in this chapter, to highlight possible areas of interaction between two particular characteristics of action domains — inertia and causality — and address the following questions:

- is there any gain in classifying fluents as inertial and non-inertial in the framework of action languages with fluent-triggered causality (fact causality) and, if yes, under what conditions;

- how does it affect possible solutions to the Ramification problem.

We believe that the best way to proceed towards this goal is to trace the evolution of action languages.

3.2 Evolution of Action Languages

As mentioned in [31], the original idea behind introducing action languages was to present a methodology enabling translations from a specialised action language to a general-purpose formalism, such as a non-monotonic reasoning system based on first-order logic. A domain described in the first action language $\mathcal{A}$ [15], for instance, can be translated [21] into a logic programming language or into the circumscriptive approach of Baker [2]. Similarly, [23] presented a translation from $\mathcal{AR}_{\mathcal{O}}$, another action language, into the formalism of nested abnormality theories [30]. In addition, subsequent research has shown that strict syntax and rigorous state transition mechanisms of action languages makes them a useful tool for understanding different aspects of reasoning about action. Ideally, "defining action languages, comparing them and studying their properties" [31] can help in "capturing our commonsense intuitions about the whole family of action domains expressible in the language" [67].

3.2.1 $\mathcal{AR}$ Languages

Following the survey of action languages presented in [48], let us consider the oldest and most developed branch of the "evolutionary tree" of action languages, that can be represented as follows:

$$\mathcal{A} \longrightarrow \mathcal{AR}^{-} \longrightarrow \mathcal{AR}_{\mathcal{O}} \longrightarrow \mathcal{AR} \longrightarrow \mathcal{ARD}.$$

The possibility of ramifications is the main feature of all "$\mathcal{AR}$"-dialects of $\mathcal{A}$ [23]. We shall treat the $\mathcal{AR}^{-}$ language [21] as the evolutionary predecessor of the language $\mathcal{AR}_{\mathcal{O}}$ [23]. Syntactically, $\mathcal{AR}^{-}$ is a subset of $\mathcal{AR}_{\mathcal{O}}$. But more importantly, the class of domains expressible in $\mathcal{AR}^{-}$ is contained in the class of domains expressible in $\mathcal{AR}_{\mathcal{O}}$. The notion of expanding expressibility of action languages and classes of domains is central to ordering the languages and constructing the tree.

All the $\mathcal{A}$-branched languages (including the original one) are capable of describing domains with deterministic actions, truth-valued fluents, inertia, and action-triggered causality. In addition, they evolve by expanding a domain's properties as shown in the following table:

	$\mathcal{A}$	$\mathcal{AR}^-$	$\mathcal{AR}_\mathcal{O}$	$\mathcal{AR}$	$\mathcal{ARD}$
dependent fluents					*
non-truth valued fluents				*	*
non-deterministic actions			*	*	*
inertial (frame) fluents		*	*	*	*
state (domain) constraints		*	*	*	*

In general, expanding the expressibility of the languages does not guarantee an enlargement of the corresponding classes of domains. Nevertheless, in the analysed evolutionary branch one can observe that such a process is at least non-diminishing:

- $\mathcal{AR}^-$ adds inertial (frame) fluents and domain constraints to $\mathcal{A}$; i.e., "$\mathcal{A}$ domains can be thought of as a special case of deterministic $\mathcal{AR}^-$ domains where there are no constraints" [22];
- $\mathcal{AR}_\mathcal{O}$ broadens the classes of deterministic and temporal projection action domains expressible in $\mathcal{AR}^-$ by introducing simple forms of non-determinism;
- $\mathcal{AR}$ [18] introduces new language elements (non-propositional fluents) but "preserves the meaning" [18] of a domain description [1];
- the contribution of $\mathcal{ARD}$ [19] is not only in showing how to represent "non-persistent ramifications" but also in introducing "a history" in the language — opening a way to formalise some of the action domains with "hypothetical reasoning".

3.2.2 $\mathcal{AC}$ Languages

The $\mathcal{AR}$-languages use state (domain) constraints to represent background knowledge. These constraints usually restrict the set of possible states and produce indirect effects of actions. All $\mathcal{AC}$-languages (named after the language $\mathcal{AC}$ introduced by Turner [67]) feature fluent-triggered (fact) causality in both these roles.

[1] In addition, the influence propositions of $\mathcal{AR}$ are supposed to correct $\mathcal{AR}_\mathcal{O}$'s handling of non-deterministic actions.

$\mathcal{AC}$ follows the approach of representing causal background knowledge that was proposed by McCain and Turner in [37] and was not formalised as an action language. Nevertheless, it is straightforward to describe an action language called, let us say, $\mathcal{AC}_{\mathcal{O}}$, based on the approach constructed in [37] that can serve as an origin of the $\mathcal{AC}$-languages branch. Formal description of $\mathcal{AC}_{\mathcal{O}}$ will be given in the next section.

As an action language, $\mathcal{AC}$ is modelled after a propositional version of the language $\mathcal{AR}$: it incorporates the idea of inertial fluents and the notion of non-determinism. In this chapter we propose to augment the basic language $\mathcal{AC}_{\mathcal{O}}$ incrementally: first, by adding the concept of a frame, and then, by extending it to non-deterministic action domains. In other words, we propose to consider yet another subset of $\mathcal{AC}$: the language $\mathcal{AC}^{-}$ that takes the intermediate place between $\mathcal{AC}_{\mathcal{O}}$ and $\mathcal{AC}$. This suggestion has a dual purpose — to present the evolution in a more systematic way and to answer the questions regarding the role of the frame concept.

The evolution along the $\mathcal{AC}$-line can be represented then in the following table:

	$\mathcal{AC}_{\mathcal{O}}$	$\mathcal{AC}^{-}$	$\mathcal{AC}$
non-deterministic actions			*
inertial (frame) fluents		*	*
fluent-triggered causality	*	*	*

The evolutionary tree of action languages is shown in Figure 3.1. Since the propositional version of $\mathcal{AR}$ differs from $\mathcal{AR}_{\mathcal{O}}$ (different meaning of "release"/ "possibly changes" propositions), the "propositional $\mathcal{AR}$" has been identified as the missing link of evolution that improves handling of non-deterministic actions compared to $\mathcal{AR}_{\mathcal{O}}$.

In addition, the tree includes a family of languages with the frame concept and fact causality introduced by Lifschitz in [31]. These languages describe deterministic action domains: the "$DL_{if} + QL^{p}$" language is aimed at temporal projection problems and the "$DL_{if} + QL^{a}$" language — at temporal explanation problems. Although these two languages allow us to represent non-propositional fluents from the moment of their appearance, it is quite natural to assume that the evolutionary tree might include their propositional counterparts as well.

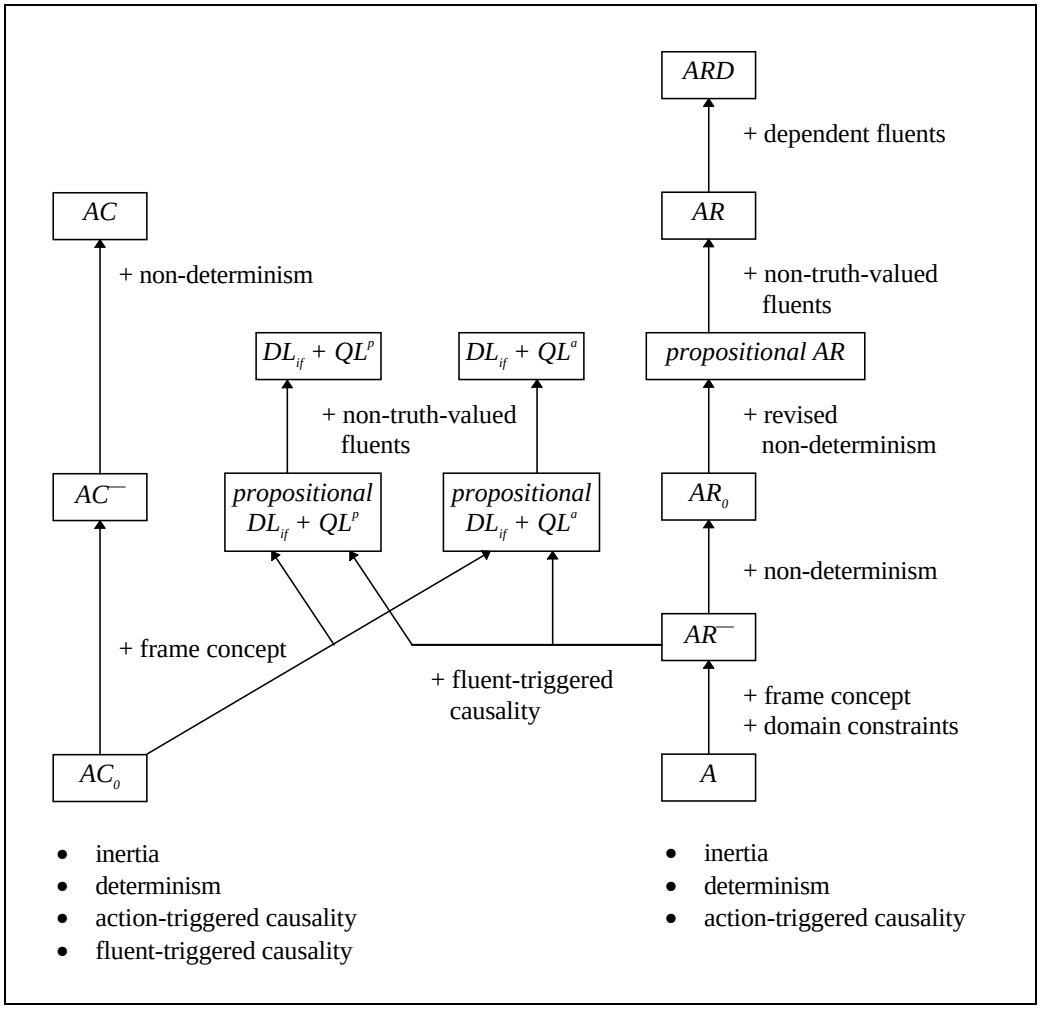

Figure 3.1: Evolutionary tree of action languages.

3.3 A Comparison between $\mathcal{AC}_{\mathcal{O}}$ and $\mathcal{AC}^{-}$ Languages

3.3.1 The $\mathcal{AC}_{\mathcal{O}}$ Language

Following [37, 67], we define the $\mathcal{AC}_{\mathcal{O}}$ language by

- a non-empty set of symbols $\mathcal{F}$, that are called fluent names, or fluents and
- a non-empty set of symbols $\mathcal{E}$, that are called action names, or actions.

In other words, the $\mathcal{AC}_{\mathcal{O}}$ language is defined by its signature $\langle \mathcal{F}, \mathcal{E} \rangle$. A fluent formula is a propositional combination of fluents. A fluent literal is an expression f or $\neg f$, where f is a fluent name. An $\mathcal{AC}_{\mathcal{O}}$ domain description consists of

- *value propositions* φ **after** Π,

  where φ is a fluent formula and Π is a string of actions. If Π is an empty string, then a value proposition is abbreviated as

  initially φ;

- *sufficiency propositions* φ **suffices for** ψ

  where φ and ψ are fluent formulae. Sometimes, a sufficiency proposition is abbreviated as φ/ψ. Another abbreviation

  always ψ

  stands for True **suffices for** ψ,

- *effect propositions* A **causes** φ **if** ψ,

  where A is an action from $\mathcal{E}$, and φ and ψ are fluent formulae.

A detailed description of the $\mathcal{AC}_{\mathcal{O}}$ language can be easily obtained from the corresponding description of the $\mathcal{AC}$ language [67]. For our purposes it is sufficient to remark that

- Ω is a set of propositions;
- $\mathcal{Q}$ is the set of all sufficiency propositions φ/ψ contained in Ω;

- for a set Λ of formulae and the set $\mathcal{Q}$ of sufficiency propositions we define the closure of a set Λ under $\mathcal{Q}$, denoted $Cn_{\mathcal{Q}}(\Lambda)$, to be the smallest set of formulae such that:

  a) $Cn_{\mathcal{Q}}(\Lambda)$ contains Λ;

  b) $Cn_{\mathcal{Q}}(\Lambda)$ is closed under classical deduction;

  c) for all $\varphi/\psi \in \mathcal{Q}$, if $\varphi \in Cn_{\mathcal{Q}}(\Lambda)$ then $\psi \in Cn_{\mathcal{Q}}(\Lambda)$;

  we write $\Lambda \mid\!\sim_{\mathcal{Q}} \varphi$ to denote that $\varphi \in Cn_{\mathcal{Q}}(\Lambda)$;

- a maximal set s of fluent literals is a state if $Cn(s)$ is closed under $\mathcal{Q}$: $Cn(s) = Cn_{\mathcal{Q}}(s)$;

- $E(s, A)$ is the set of fluent formulae φ for which there is an effect proposition

  $$A \textbf{ causes } \varphi \textbf{ if } \psi$$

  in Ω such that s satisfies ψ.

We say that an action A is prohibited in a state s if there is an effect proposition

$$A \textbf{ causes } \text{False} \textbf{ if } \psi$$

in Ω such that s satisfies ψ. Consequently, a set $E = E(s, A)$ is an explicit effect of A at s if A is not prohibited in s.

Finally, a state s' may result from doing A in s if there is an explicit effect E of A in s such that

$$Cn(s') = Cn_{\mathcal{Q}}((s \cap s') \cup E) \tag{3.1}$$

Intuitively, a state s' specified by this condition satisfies explicit effect E and contains as few changes with respect to s as possibly justifiable by the underlying causal "laws" (sufficiency propositions) in $\mathcal{Q}$. In other words, any change from s to s' must be "explained" by causal laws applied (as inference rules) to explicit effects and unchanged literals (common to both s and s'). Formally, we can define a state transition (selection) mechanism of the causal action language $\mathcal{AC}_{\mathcal{O}}$, based on the following selection function:

$$Res_{\mathcal{AC}_{\mathcal{O}}}(s, E) = \{s' : Cn(s') = Cn_{\mathcal{Q}}((s \cap s') \cup E)\}.$$

It is easy to see that this fixed-point definition is equivalent to the definition of successor states $Res_\mathcal{Q}(s, E)$ given by McCain and Turner in [37], where $\mathcal{Q}$ is the causal system (a set of causal inference rules or "laws"), and allows us to obtain the same successor states as McCain and Turner's approach. In other words, the $\mathcal{AC}_\mathcal{O}$ language (more precisely, its state transition mechanism) is selection-equivalent to causal systems specified in [37]. Formally,

$$Res_\mathcal{Q}(s, E) = Res_{\mathcal{AC}_\mathcal{O}}(s, E).$$

We often refer to the elements of $Res_\mathcal{Q}(s, E) = Res_{\mathcal{AC}_\mathcal{O}}(s, E)$ as *causal fixed-points*.

In describing the semantics of the $\mathcal{AC}_\mathcal{O}$ language, we shall follow [67]. A set Π of action strings is *prefix-closed* if, for every string $\pi \in \Pi$, every prefix of π is a also in Π. A *structure* for a domain description Ω is a partial function from action strings to interpretations of the fluent atoms, whose domain is non-empty and prefix-closed. The domain of a structure Ψ is denoted as $Dom(\Psi)$. For every structure Ψ, $Dom(\Psi)$ includes the empty string (denoted by α). Given a structure Ψ, it is said that an atomic value proposition φ **after** Π is $true$ in Ψ if $\Pi \in Dom(\Psi)$ and $\Psi(\Pi)$ satisfies φ. Then the general truth definition for value propositions is given by standard recursion over the propositional connectives.

A structure Ψ for Ω is called a *model* of Ω if it satisfies the following four conditions:

- $\Psi(\alpha)$ is a state;
- for all action strings $\Pi \in Dom(\Psi)$ and all actions A, if $Res_{\mathcal{AC}_\mathcal{O}}(A, \Psi(\Pi)) \neq \emptyset$ then the string of actions $\Pi; A$ is also in $Dom(\Psi)$;
- for all strings $\Pi; A \in Dom(\Psi)$, $\Psi(\Pi; A) \in Res_{\mathcal{AC}_\mathcal{O}}(A, \Psi(\Pi))$;
- every value proposition in Ω is true in Ψ.

3.3.2 The $\mathcal{AC}^-$ Language

Since there is no "frame/non-frame distinction" in the causal action language $\mathcal{AC}_\mathcal{O}$, all fluents are subject to inertia. The language DL_{if} was introduced in [31] to overcome

this particular limitation. However, as we intend to show, the categorisation of fluents according to their inertial properties does not really refine selection functions of causal action languages.

Let us follow the evolution illustrated in Figure 3.1. In order to introduce $\mathcal{AC}^-$, we need to augment $\mathcal{AC}_\mathcal{O}$ with the concept of a frame and appropriately change the fixed-point condition 3.1.

To achieve this we designate the elements of a certain subset, Φ, of the set $\mathcal{F}$ as inertial. The set Φ is called a frame designation for a domain description Ω if it is not empty, and is included in the augmented language signature $\langle \mathcal{F}, \Phi, \mathcal{E} \rangle$. The signature $\langle \mathcal{F}, \mathcal{E} \rangle$ of the $\mathcal{AC}_\mathcal{O}$ language is subsumed by the signature of the $\mathcal{AC}^-$ language and can be considered as $\langle \mathcal{F}, \emptyset, \mathcal{E} \rangle$.

Let L_Φ denote the set of inertial fluent literals, where a literal (f or $\neg f$) is inertial if a fluent name f occurring in it is inertial ($f \in \Phi$). A fluent formula is called inertial if its fluent atoms are inertial. The following fixed-point condition defines a state s' that may result according to the $\mathcal{AC}^-$ state selection (transition) mechanism from applying A at s,

$$Cn(s') = Cn_\mathcal{Q}((s \cap s' \cap L_\Phi) \cup E) \tag{3.2}$$

where E is an explicit effect of A in s as before. Again, a state s' specified by this condition satisfies explicit effect E and contains as few changes with respect to s as possibly justifiable by the underlying causal rules in $\mathcal{Q}$. However, unlike condition 3.1, this condition allows only *inertial literals* common to both states s and s' to contribute (together with explicit effects) to causal inference driven by rules in $\mathcal{Q}$. The following selection function captures the state transition mechanism of the $\mathcal{AC}^-$ language:

$$Res_{\mathcal{AC}^-}(s, E) = \{s' : Cn(s') = Cn_\mathcal{Q}((s \cap s' \cap L_\Phi) \cup E)\}.$$

It is interesting to compare at this point condition 3.2 with the following similar condition

$$s' \cap L_\Phi = Cn_\mathcal{Q}((s \cap s' \cap L_\Phi) \cup E) \cap L_\Phi, \tag{3.3}$$

which basically defines a state transition mechanism for the language DL_{if} [31]. In this case, the Principle of Minimal Change is applied only to inertial literals — in other

words, a successor state must contain as few changes as possible *in terms of inertial literals*, and the changes in non-inertial literals are disregarded. Similarly to the previous case, all the "important" changes should be causally justified by explicit effects and/or *inertial literals* common to both initial and successor states. The following selection function specifies the state transition mechanism of the DL_{if} language:

$$Res_{DL_{if}}(s, E) = \{s' : s' \cap L_\Phi = Cn_{\mathcal{Q}}((s \cap s' \cap L_\Phi) \cup E) \cap L_\Phi\}.$$

Condition 3.3 defines successor states that, in general, do not necessarily satisfy condition 3.2. The reason is that there are fewer changes to account for, if only inertial literals are regarded as important and therefore a state s' selected as a successor state according to condition 3.3, may still have some non-inertial literals left unsupported. On the other hand, a state s' satisfying 3.2 should, obviously, satisfy 3.3 — as all changes (in inertial literals or otherwise) are justified. Formally,

$$Res_{\mathcal{AC}^-}(s, E) \subseteq Res_{DL_{if}}(s, E).$$

However, these two selection functions become equivalent if we require that an $\mathcal{AC}^{-}$ domain description includes an explicit definition

$$\textbf{always } f \equiv \varphi$$

where φ is an inertial fluent formula[2], for each non-inertial fluent f. In this case, if every inertial literal in a successor state is justified, then a change in every non-inertial literal is explained as well, following from explicit definitions.

Moreover, in [67] Turner indicated that under the requirement of explicitly defining every non-inertial fluent in a domain, the domain description Ω becomes both an $\mathcal{AC}$ domain description and a propositional $\mathcal{AR}$ domain description, and the $\mathcal{AC}$ models of Ω are exactly the propositional $\mathcal{AR}$ models of Ω. In other words, the $\mathcal{AR}$ Theorem [67] establishes a "horizontal" connection between different branches of the evolutionary tree, proving selection-equivalence between respective state transition mechanisms.

We shall now attempt to identify conditions (more precisely, restrictions) which ensure selection-equivalence and preserve meaning of domain descriptions produced in

[2] An inertial fluent formula is a formula whose atoms are inertial fluents.

the $\mathcal{AC}_{\mathcal{O}}$ language, when moving "vertically" to the $\mathcal{AC}^{-}$ language and employing the frame concept. Relaxing such restrictions will demonstrate what is gained by categorising fluents as inertial and non-inertial in the framework of action languages with fluent-triggered causality (fact causality).

3.3.3 A Connection between $\mathcal{AC}_{\mathcal{O}}$ and $\mathcal{AC}^{-}$ Languages

As mentioned in [18], domain constraints in the $\mathcal{AR}$ language play two different roles: they (globally) affect selection of possible successor states, and they also (locally) determine indirect effects of actions. Like the DL_{if} language, the $\mathcal{AC}^{-}$ language does not impose the requirement to explicitly define every non-inertial fluent in its domain description. This leaves some of the non-inertial fluents less supported as far as indirect effects are concerned. In other words, the "ramification constraints" [37, 67] expressed in the form of sufficiency propositions "φ **suffices for** ψ" become the only source for updating the truth values of non-inertial fluents without explicit definitions, when an action is performed. Here we introduce an analogue of the explicit definition adopted in the $\mathcal{AR}$ and the $\mathcal{AC}$ languages.

Definition 3.3.1 *(Effect-complete fluent)*
A fluent f is called an effect-complete fluent in an $\mathcal{AC}^{-}$ (or $\mathcal{AC}$) domain description Ω, if and only if Ω contains propositions

$$\varphi \textbf{ suffices for } f,$$
$$\neg\varphi \textbf{ suffices for } \neg f,$$

for some inertial formula φ.

We will denote the set of all effect-complete fluents by Σ, and the set of all effect-complete literals by L_{Σ}. The following observation establishes a connection between the languages $\mathcal{AC}_{\mathcal{O}}$ and $\mathcal{AC}^{-}$.

Lemma 3.3.2 *Let Ω^{-} be an $\mathcal{AC}^{-}$ domain description such that each non-inertial fluent is an effect-complete fluent:*

$$\mathcal{F} \setminus \Phi \subseteq \Sigma.$$

Then $\Omega_0 = \Omega^-$ defined in the signature $\langle\, \mathcal{F}, \emptyset, \mathcal{E} \,\rangle$ by abandoning the frame designation Φ is an $\mathcal{AC}_\mathcal{O}$ domain description, and the $\mathcal{AC}_\mathcal{O}$ models of Ω_0 are exactly the $\mathcal{AC}^-$ models of Ω^-.

In other words, if each non-inertial fluent is an effect-complete fluent, then the $\mathcal{AC}_\mathcal{O}$ state transition mechanism is selection-equivalent to the $\mathcal{AC}^-$ state transition mechanism:

$$Res_{\mathcal{AC}_\mathcal{O}}(s, A) = Res_{\mathcal{AC}^-}(s, A).$$

This means that not much can be really gained by supplying a domain with a frame designation, while extending from $\mathcal{AC}_\mathcal{O}$ to $\mathcal{AC}^-$, if all non-inertial fluents are effect-complete. In fact, in this case (and in a similar case, when all non-inertial fluents are explicitly defined) the non-inertial fluents are nothing but syntactic abbreviations for some inertial formulae. If, however, these restrictions (binding non-inertial fluents) are lifted, state transitions may result in somewhat unexpected state selections.

To illustrate this, we consider an example enhancing the Two-Switches example [23, 37, 67]. In particular, we aim to demonstrate differences among $\mathcal{AC}_\mathcal{O}$, $\mathcal{AC}^-$, and DL_{if} languages, influenced by inter-relations between the sets Σ and $\mathcal{F} \setminus \Phi$.

Example 3.3.3 *(Light and Shadows)*
There are two switches, a light, and shadows. The light is on only when both switches are on. Light "causes" shadows to appear or, in other words, light is "sufficient" for the appearance of shadows. The $\mathcal{AC}_\mathcal{O}$ domain can be described using four propositional fluent names:

$$\mathcal{F} = \{Switch_1, Switch_2, Light, Shadows\}.$$

There are two action names:

$$\mathcal{E} = \{Toggle_1, Toggle_2\}.$$

The propositions are:

$$Toggle_1 \textbf{ causes } \neg Switch_1 \textbf{ if } Switch_1, \tag{3.4}$$

$$Toggle_1 \textbf{ causes } Switch_1 \textbf{ if } \neg Switch_1, \tag{3.5}$$

$$Toggle_2 \textbf{ causes } \neg Switch_2 \textbf{ if } Switch_2, \tag{3.6}$$

$$Toggle_2 \textbf{ causes } Switch_2 \textbf{ if } \neg Switch_2, \tag{3.7}$$

$$Switch_1 \wedge Switch_2 \textbf{ suffices for } \textit{Light}, \tag{3.8}$$

$$\neg Switch_1 \vee \neg Switch_2 \textbf{ suffices for } \neg \textit{Light}, \tag{3.9}$$

$$\textit{Light} \textbf{ suffices for } \textit{Shadows}, \tag{3.10}$$

$$\textbf{initially } \neg Switch_1, \tag{3.11}$$

$$\textbf{initially } \neg Switch_2, \tag{3.12}$$

$$\textbf{initially } \neg \textit{Light}, \tag{3.13}$$

$$\textbf{initially } \neg \textit{Shadows}. \tag{3.14}$$

Consider the action $Toggle_2$ performed in the initial state

$$s = \{\neg Switch_1, \neg Switch_2, \neg Light, \neg Shadows\}.$$

It is easy to check that the only possible next state satisfying condition 3.1 is

$$s_1 = \{\neg Switch_1, Switch_2, \neg Light, \neg Shadows\}$$

and this, indeed, is the intuitive choice. In other words, the $\mathcal{AC}_{\mathcal{O}}$ selection mechanism handles the action well:

$$Res_{\mathcal{AC}_{\mathcal{O}}}(s, Toggle_2) = \{s_1\}.$$

Now let us consider the $\mathcal{AC}^-$ domain description based on 3.4 — 3.14 with at least one fluent which is non effect-complete and non-inertial. The frame designation

$$\Phi = \{Switch_1, Switch_2\} \tag{3.15}$$

leaves the fluent $Shadows$ non-inertial, and propositions 3.4—3.14 do not define it as effect-complete, unlike the fluent $Light$ which is non-inertial but effect-complete.

It is worth noting that the $\mathcal{AC}^-$ domain description 3.4—3.15 is not a "legal" $\mathcal{AC}$ domain description because the non-inertial fluent $Shadows$ is not explicitly defined in

terms of inertial fluents. However, this description is a "legal" one in the $DL_{if} + QL^p$ language (if we rename all axioms **initially** φ to **now** φ).

Again consider the action $Toggle_2$ performed in the initial state s. It turns out that there are no interpretations satisfying condition 3.2. The task to derive a value of the non-inertial fluent $Shadows$ fails because this fluent is neither effect-complete nor explicitly defined in terms of inertial fluents. In other words, the state selection mechanism of the $\mathcal{AC}^-$ language produces no possible next states (and no models) for the action:

$$Res_{\mathcal{AC}^-}(s, Toggle_2) = \emptyset.$$

The $DL_{if} + QL^p$ domain description 3.4—3.15, however, treats two next states as possible:

$$s_1 = \{\neg Switch_1, Switch_2, \neg Light, \neg Shadows\}$$

$$s_2 = \{\neg Switch_1, Switch_2, \neg Light, Shadows\},$$

each of which satisfies condition 3.3. In other words, the state selection mechanism of the DL_{if} language can not decide on the value of the non-inertial fluent $Shadows$, disregarding the corresponding change:

$$Res_{DL_{if}}(s, Toggle_2) = \{s_1, s_2\}.$$

3.3.4 Discussion

Disagreement among possible successor states (and therefore, models) obtained by the $\mathcal{AC}_\mathcal{O}$, the $\mathcal{AC}^-$ and the DL_{if} languages can be avoided either by a better frame designation of the domain description including the non effect-complete fluent $Shadows$ in the frame, or by the introduction of a cause leading to the disappearance of "shadows", for example,

$$\neg Light \textbf{ suffices for } \neg Shadows.$$

Consider the first way (a better frame designation) for our example. Let

$$\Phi = \{Switch_1, Switch_2, Shadows\} \tag{3.16}$$

be the new frame designation. Then the domain description 3.4—3.14, 3.16 entails

$$Res_{\mathcal{AC}^-}(s, Toggle_2) = Res_{DL_{if}}(s, Toggle_2) = \{s_1\},$$

where s_1 is a unique successor state according to both the $\mathcal{AC}^-$ and the DL_{if} state selection mechanisms.

However, attachment of the non effect-complete fluent $Shadows$ to the frame does not make reasoning about domain actions more intuitive. For instance, consider a case when the action $Toggle_2$ is performed in the state

$$s' = \{Switch_1, Switch_2, Light, Shadows\}.$$

Now all three languages ($\mathcal{AC}_\mathcal{O}$, $\mathcal{AC}^-$ and DL_{if}) agree that the only next possible state must be

$$s'' = \{Switch_1, \neg Switch_2, \neg Light, Shadows\}.$$

More precisely:

$$Res_{\mathcal{AC}_\mathcal{O}}(s', Toggle_2) = Res_{\mathcal{AC}^-}(s', Toggle_2) = Res_{DL_{if}}(s', Toggle_2) = \{s''\}.$$

The fluent $Shadows$ belongs to the frame and, therefore, keeps its value through the change from s' to s''. Since no one indicated a sufficient reason for "shadows" to disappear, it behaves like a persistent ramification (of some previous action), and stays even when the "light" is switched off. This does not, probably, contradict our intuition greatly (maybe there are other reasons for "shadows" to remain), but a designation of the fluent $Shadows$ as a non-inertial fluent would be, perhaps, more intuitive. However, as example 3.3.3 illustrated, this did not yield satisfactory state selections. In short, attachment of this fluent to the frame is necessitated by a state transition mechanics, and not by domain knowledge.

The second way (introducing a cause for the disappearance of "shadows") is not always available because, in general, an agent performs reasoning about action in the absence of complete information. Besides, if we necessarily require from each non-inertial fluent to be effect-complete (or explicitly defined, for that matter), then the provided solution to the Ramification problem does not look like a principled one. As mentioned

earlier, in this case the non-inertial fluents serve as syntactic abbreviations of some inertial formulae.

Instead, one may be tempted to refine a state transition system and consider yet another action language. Let us define a new language, labeled $\mathcal{AC}^{*}$, exactly as $\mathcal{AC}^{-}$, except that the state transition system is given by the following two-tiered selection function

$$Res_{\mathcal{AC}^{*}}(s,E) = \{s' \in Res_{DL_{if}}(s,E) : \\ s' \cap (L_{\mathcal{F}} \setminus L_{\Phi}) = Cn_{\mathcal{Q}}((s \cap s') \cup E) \cap (L_{\mathcal{F}} \setminus L_{\Phi})\}, \tag{3.17}$$

where

$$Res_{DL_{if}}(s,E) = \{s' : s' \cap L_{\Phi} = Cn_{\mathcal{Q}}((s \cap s' \cap L_{\Phi}) \cup E) \cap L_{\Phi}\}, \tag{3.18}$$

and $L_{\mathcal{F}}$ is the set of all fluent literals. In other words, the state transition system 3.17—3.18 works in two steps:

- firstly, it selects the states satisfying the fixed-point condition 3.3 (which is identical to the condition used in 3.18) — thus agreeing with the state selection mechanism of the DL_{if} language and minimising changes only in inertial literals L_{Φ};
- secondly, among states selected in the first step, it minimises changes in non-inertial literals ($L_{\mathcal{F}} \setminus L_{\Phi}$).

It is easy to see that

$$Res_{\mathcal{AC}^{*}}(s,E) \subseteq Res_{DL_{if}}(s,E),$$

so the state selection mechanism of the $\mathcal{AC}^{*}$ language would never prefer more states than the state selection mechanism of the DL_{if} language. For instance, revisiting example 3.3.3 with two switches in the frame, if the action $Toggle_2$ is again performed in the initial state s then the state

$$s_1 = \{\neg Switch_1, Switch_2, \neg Light, \neg Shadows\}$$

is the only member of $Res_{\mathcal{AC}^{*}}(s, Toggle_2)$, being "short-listed" from

$$Res_{DL_{if}}(s, Toggle_2) = \{s_1, s_2\}$$

as the state with less changes in non-inertial literals.

However, this approach does not seem to be promising enough. It is not evident that more complex domains will not require more "layers" of minimisation applied to fluents with varied degrees of inertia. Moreover, in some domains it can be difficult to properly categorise fluents. For instance, it has been shown (the relay example [62]) that it does not appear possible to select only intended resulting states according to the state transition (selection) mechanism of the $\mathcal{AR}_{\mathcal{O}}$ language for any categorisation of fluents as inertial and non-inertial. As an alternative, another method was proposed — the causal relationship approach of Thielscher [62]. It is worth noting, however, that the languages considered with fluent-triggered causality and the frame concept ($\mathcal{AC}^{-}$, DL_{if} and $\mathcal{AC}^{*}$) handle the relay example as intended, provided that all non-inertial fluents are chosen from effect-complete ones. But the simpler $\mathcal{AC}_{\mathcal{O}}$ language also successfully deals with the relay example — and does it without employing fluent categorisation.

3.4 Summary

The idea of a compact and concise representation dates back at least as far as Occam's Razor: other things being equal, simple theories are to be preferred to complex ones. Employment of the frame concept (the policy of categorisation) in action languages operating with causality might have helped to advance us towards a more concise solution to the Frame and the Ramification problems. Unfortunately, the solutions considered in this chapter impose, from our point of view, too severe restrictions on a domain description: each non-inertial fluent has to be defined in terms of inertial fluents — either through an explicit definition, or as an effect-complete fluent. This demand decreases the value of the frame concept in the framework of languages with fluent-triggered causality — at least, in the considered branch of the evolutionary tree. Probably, further development of action languages will clarify what role the policy of categorisation (and the frame concept, in particular) should play in logic-based approaches to reasoning about action, change and causality.

However, at this stage we shall concentrate on our primary goal — providing a unifying semantics for action theories based on the Principles of Minimal and Causal change.

To achieve this goal, we shift our focus from the action languages framework to the original approach that motivated development of the $\mathcal{AC}$ branch — McCain and Turner's action theory with causal fixed-points. This framework combines the principle of minimal change with that of causal change, and our intention is to furnish a generic preferential style semantics for the state selection mechanism of causal fixed-points. This semantics would not involve a policy of categorisation and, therefore, could advance us towards a unification with the causal relationship approach of Thielscher, that was proposed as a superior solution to the Frame and Ramification problems [62].

Chapter 4

Causal Systems with Fixed-Points

In this chapter we examine the causal theory of actions put forward by McCain and Turner [37] for determining ramifications. Our principal aim is to provide a characterisation of this causal theory of actions in terms of the augmented preferential semantics. This would allow us to compare it with other logics of action and highlight the nature of context-sensitive causality underlying their proposal. We begin by showing that our goal is not achievable by a preferential mechanism alone. At this point we do not abandon preferential semantics altogether but augment it in order to arrive at the desired result — in the spirit of our semantics introduced in Chapter 2. Thus, we demonstrate that two components may be required to provide a concise solution to the Frame and Ramification problems: minimal change under a preferential structure and context-sensitive causality.

4.1 Background

As mentioned in previous chapters, preferential semantics remain an important and useful concept. Intuitively, a preferential structure on possible states of the world may represent (from the reasoning agent's point of view) how plausible one state is with respect to another, or how similar one state is with another, or how easy (in terms of effort, resources) it is to reach one state from another, etc. Only the most preferred (most plausible, most similar, most accessible) states are to be considered as serious possibilities according to a preferential-style semantics. In other words, the *Principle of Minimal Change* suggests a bias towards the minimal models when there are multiple successor

states satisfying the direct effects of an action.

Informally, one may say that a preferential structure sets some soft constraints on possible states, and the agent should try to satisfy these constraints as long as ordinary domain constraints are not violated (these constraints divide the state-space into legitimate and illegitimate states).

However, it has been recognised recently that traditional domain constraints alone are not sufficient to provide compact solutions to the Frame and Ramification problems [29, 32, 33, 37, 62, 3]. Consequently, the notion of causality has been considered in the literature [33, 37, 63, 56] as an (additional) element required for reasoning about action and change. The causal constraints were introduced instead of (or in addition to) domain constraints [33, 37, 38, 62, 68], and the Principle of Causal Change was suggested as a new guide, favouring successor states where changes are necessitated (caused).

Let us consider, for instance, a variant of the well-known Two-Switches example, introduced by Lifschitz [29][1] and represented as the Two-Locks example by Lin [33]. There are three fluents in this domain:

$$\{Switch_1, Switch_2, Light\},$$

and one domain constraint

$$Switch_1 \wedge Switch_2 \leftrightarrow Light.$$

Intuitively, the constraint is meant to indicate that the light is on if and only if both switches are up. Let us now consider the initial state

$$\{\neg Switch_1, Switch_2, \neg Light\},$$

and perform the action of toggling the first switch with post-condition $Switch_1$. A system with a pure preferential semantics (based, for instance, on symmetric difference) may produce two successor states [37]:

$$\{Switch_1, Switch_2, Light\},$$

$$\{Switch_1, \neg Switch_2, \neg Light\},$$

[1] This work revealed the insufficiency of domain constraints for solving the Ramification Problem.

each of which differs from the initial one strictly in two literals. The second successor state is, however, counter-intuitive — why should the second switch be affected? This unintended ramification appears because the original constraint is not strong enough, and in fact, logically implies the following:

$$Switch_1 \wedge \neg Light \supset \neg Switch_2.$$

To faithfully represent the intended ramification something stronger than the original constraint is needed, as pointed out by Lin [33]. In the considered example the fact that both of the switches are in the up position *causes* the light to be on, and the fact that one of the switches is not up *causes* the light to be off.

Among many interesting proposals addressing the question of how to operate with causal statements like this one, the causal theory of actions put forward by McCain and Turner [37] in 1995, is quite prominent. It has influenced the area of causal reasoning about action for the last few years — several action languages appeared in response to this proposal [31, 67], and a number of parallel and competing approaches used it as a benchmark [63, 20]. At the same time, the proposal was clear and included causal components rather elegantly.

McCain and Turner introduce causal laws (rules) of the form $\phi \Rightarrow \psi$, where ϕ and ψ are fluent formulae (i.e., they do not contain further instances of $\Rightarrow$ but only classical truth functional connectives). Intuitively, these formulae can be read as 'ϕ *causes* ψ', and express "a relation of determination between states of affairs that make ϕ and ψ true" [37]. In fact, traditional domain constraints can be subsumed by causal laws [37, Proposition 3]. An important point to notice is that causal laws behave as 'unidirectional' implications — the contrapositive ($\neg\psi \Rightarrow \neg\phi$) does not hold in general. Revisiting the Two-Switches example, we note that the following causal rules would ensure the intended ramifications:

$$Switch_1 \wedge Switch_2 \Rightarrow Light,$$

$$\neg Switch_1 \vee \neg Switch_2 \Rightarrow \neg Light.$$

In short, the proposal of McCain and Turner [37] describes a causal theory of actions, where changes in state variables are intended to be minimal, subject to satisfaction of all

causal laws. In other words, both Principles of Minimal and Causal Change are put to work.

We show first that it is not possible to characterise McCain and Turner's causal theory via a pure preferential semantics applied to interpretations of the original (unaugmented) language[2]. We then augment preferential semantics with a relational structure, while staying within the approximation $\mathcal{W} = \Gamma$.

This is done in two steps, each of which describes a particular state-selection mechanism: firstly, *state elimination systems* and secondly, *state transition systems*. State elimination systems employ a preference structure based on symmetric difference (the PMA ordering) and an auxiliary binary relation on *sets of states*. State transition systems use a binary relation on *states*, constructed from the auxiliary relation used in elimination systems. Thus, state transition systems can be described with a variant of our augmented preferential semantics for logics of actions operating with causality, introduced in Chapter 2[3].

In summary, the obtained results allow us to identify the character of context-sensitive causality exploited in McCain and Turner's causal theory of actions, and show that minimal change and causality can co-exist in separate roles and complement each other.

This chapter is structured as follows. In the next section we shall sketch some basic terminology and notation followed by an overview of McCain and Turner's [37] causal theory of action. In section 4.3 we shall show that it is not possible to supply a straightforward preferential semantics to capture McCain and Turner's approach. The solution we suggest here is not to abandon preferential semantics entirely but rather to augment it. In sections 4.4 — 4.6 we investigate the different state selection mechanisms. We end with a discussion of the significance of these results.

[2] A similar result was sought by Peppas et al. [45] but the counterexample they present assumes a transitive and total ordering where we only assume transitivity.

[3] The semantics developed by Peppas et al. [45] is similar but characterises only a subset of the possible McCain and Turner causal systems whereas the semantics we present captures all.

4.2 Causal Systems

In this section we review McCain and Turner's [37] causal theory of actions, and reproduce, for convenience, the technical preliminaries described in Chapter 2. Throughout this chapter we shall be working with a fixed finitary propositional language $\mathcal{B}$ whose propositional letters we shall call *fluents*. The set of all fluents is denoted by $\mathcal{F}$. A *literal* is a fluent or the negation of a fluent. A *state* (or *world*) is defined as a maximal consistent set of literals. The set of all literals will be denoted $\mathcal{N}$. The set of all states will be denoted $\mathcal{W}$. By $[\phi]$ we denote all states consistent with the sentence $\phi \in \mathcal{B}$ (i.e., $[\phi] = \{w \in \mathcal{W} : w \vdash \phi\}$). Occasionally we will refer to $[\phi]$ as the ϕ-*states* (or ϕ-*worlds*).

As outlined above, McCain and Turner introduce a new connective $\Rightarrow$ to denote a causal relationship between sentences ϕ and ψ of the underlying language $\mathcal{B}$. This allows for expressions of the form $\phi \Rightarrow \psi$ (where $\phi,\ \psi \in \mathcal{B}$) which are termed *causal laws* (or *causal rules*). Nesting of $\Rightarrow$ is not permitted. For the sake of simplicity we shall assume here that the antecedent of any causal law is consistent. A set of causal laws $\mathcal{Q}$ is referred to as a *causal system*. Given any set of sentences $\Lambda \subseteq \mathcal{B}$ and a causal system $\mathcal{Q}$, the (causal) *closure* of Λ in $\mathcal{Q}$ is denoted $C_{\mathcal{Q}}(\Lambda)$ and defined to be the smallest superset of Λ closed under classical logical consequence and such that for any $\phi \Rightarrow \psi \in \mathcal{Q}$, if $\phi \in C_{\mathcal{Q}}(\Lambda)$, then $\psi \in C_{\mathcal{Q}}(\Lambda)$. We also say that Λ *causally implies* ϕ with respect to $\mathcal{Q}$ if and only if $\phi \in C_{\mathcal{Q}}(\Lambda)$ and denote this by $\Lambda \mid\!\sim_{\mathcal{Q}} \phi$.

Another notion that will be of importance is that of a *legitimate state* with respect to a causal system $\mathcal{Q}$. Any state r is legitimate with respect to $\mathcal{Q}$ if and only if $r = C_{\mathcal{Q}}(r) \cap \mathcal{N}$. That is, a state is legitimate if and only if it does not contravene any causal laws of $\mathcal{Q}$. The set of legitimate states with respect to $\mathcal{Q}$ is denoted by $\mathcal{W}_{\mathcal{Q}}$.

In McCain and Turner's framework, actions are referred to only through their direct effects (post-conditions). McCain and Turner's aim is to determine the set of possible next (or resultant) states $Res_{\mathcal{Q}}(w,\ E)$ given an initial state w and the direct effects (or post-conditions) of an action represented by the sentence E. Formally speaking, we have for any causal system $\mathcal{Q}$ a function $Res_{\mathcal{Q}}$ mapping a legitimate (initial) state w and sentence E (direct effects) to the set of states $Res_{\mathcal{Q}}(w,\ E)$ according to the definition

[37]:

$$r \in Res_{\mathcal{Q}}(w,\ E) \text{ if and only if } r = \{\mathtt{f} \in \mathcal{N} :\ (w \cap r) \cup \{E\} \mid\!\sim_{\mathcal{Q}} \mathtt{f}\}$$

Intuitively speaking, the elements of $Res_{\mathcal{Q}}(w,\ E)$ are simply those E-states where all changes with respect to w can be justified by the underlying causal system. We often refer to the elements of $Res_{\mathcal{Q}}(w,\ E)$ as *causal fixed-points*. Note that it follows from this definition that if $r \in Res_{\mathcal{Q}}(w,\ E)$, then $r \in [E]$ (i.e., r must satisfy the direct effects of the action).

It would be useful to consider, at this stage, a few examples illustrating selection of successor states according to the selection mechanism $Res_{\mathcal{Q}}(w,\ E)$. Of particular interest, is a comparison with successor states obtained by an action system where all causal rules $\phi \Rightarrow \psi$ are simply replaced by material implications $\phi \supset \psi$. The differences would highlight the emphasis of the causal fixed-points approach on the unidirectional flow of causality.

Let us consider a simple system (or domain) with two fluents a and b, a causal constraint $a \Rightarrow b$, and two actions — one with the post-condition a, and another with the post-condition $\neg b$. We shall compare this domain with another, where the causal constraint $a \Rightarrow b$ is replaced with the domain constraint $a \supset b$.

First of all, we should point out that these two "parallel" domains share the same state-space $\{\{a, b\}, \{a, \neg b\}, \{\neg a, b\}, \{\neg a, \neg b\}\}$, and have exactly the same legitimate states $\mathcal{D}$, excluding the state $\{a, \neg b\}$ because it violates the constraint ($a \Rightarrow b$ in the first domain, and $a \supset b$ in the second).

Let the initial state w be $\{\neg a, \neg b\}$. By the definition,

$$Res_{\mathcal{Q}}(w, a) = \{\{a, b\}\}.$$

In the parallel domain (where the causal constraint is replaced with a material implication), we may employ a preferential structure based, for convenience, on the PMA ordering $\prec$ and obtain

$$Res_0(w, a) = min(\prec_w, [a]) \cap \mathcal{D} = \emptyset,$$

and

$$Res_1(w, a) = min(\prec_w, \mathcal{D} \cap [a]) = \{\{a, b\}\}.$$

The function $Res_0(w, a)$ selected the state $\{a, \neg b\}$ as the only $\prec_w$-minimal element in $[a]$, and then ruled it out as illegitimate. The function $Res_1(w, a)$, on the other hand, "went beyond" the (empty) first boundary, and selected the minimal element in the set $\mathcal{D} \cap [a]$ — the state $\{a, b\}$ or the only state lying on the second boundary. In other words, the stronger $Res_0(w, a)$ turned out to be too demanding compared with $Res_{\mathcal{Q}}(w, a)$, while the weaker $Res_1(w, a)$ was precise in characterising causal fixed-points.

Now let us treat the state $\{a, b\}$ as the initial state, say w', and consider the second action with the post-condition $\neg b$. The causal fixed-points approach yields

$$Res_{\mathcal{Q}}(w', \neg b) = \emptyset,$$

ruling out the state $\{\neg a, \neg b\}$ (the only legitimate possibility among the post-conditions states $[\neg b]$), because $\neg b$ could not causally justify $\neg a$.

In the parallel domain, we obtain the following:

$$Res_0(w', \neg b) = min(\prec_{w'}, [\neg b]) \cap \mathcal{D} = \emptyset,$$

and

$$Res_1(w', \neg b) = min(\prec_{w'}, \mathcal{D} \cap [\neg b]) = \{\{\neg a, \neg b\}\}.$$

This time, the stronger $Res_0(w, a)$ (reaching only the first boundary) was precise in characterising causal fixed-points, while the weaker $Res_1(w, a)$ (reaching the second boundary) was not demanding enough (note that the domain constraint $a \supset b$ has a contrapositive $\neg b \supset \neg a$ unlike the causal rule).

This example clearly shows that, although in general,

$$Res_0(w, E) \subseteq Res_{\mathcal{Q}}(w, E) \subseteq Res_1(w, E),$$

neither preferential selection function is precise in characterising causal fixed-points correctly.

We are now in a position to state our aims more clearly. The desire is to mimic McCain and Turner's fixed-point definition using a preference ordering over states and in such a way as not to introduce auxiliary sentences into our language. More specifically, we wish to investigate whether this is at all possible; whether we can provide a preferential-style semantics characterising $Res_{\mathcal{Q}}(w,\ E)$ for any legitimate state w and sentence E.

4.3 Impossibility Results

In this section we clearly specify what we mean by a preferential semantics. We then present an impossibility result showing that a traditional preferential semantics is not capable of characterising McCain and Turner's fixed-point definition.

We are given an initial state $w \in \mathcal{W}$ and a (strict) preference ordering $<_w \subseteq \mathcal{W} \times \mathcal{W}$ on states. The only restriction we place on $<_w$ here is that it satisfy transitivity. Adhering to the essence of preferential semantics [60] we seek to define those states resulting from the occurrence of an action with direct effects E at initial state w as the minimal E-states under $<_w$. The following condition expresses these desiderata:

(P) $Res_{\mathcal{Q}}(w,\ E) = min(<_w,\ [E])$

Now, according to McCain and Turner, there is no need to consider illegitimate states as possible resultant states since they contradict the causal laws. Hence, we begin by focusing on a variant of condition (P):

(P′) $Res_{\mathcal{Q}}(w,\ E) = min(<_w,\ [E]) \cap \mathcal{D}$

We are now in a position to state a fundamental result of this section; that, in general, it is not possible to satisfy the condition (P′) (with transitive $<_w$). Note firstly that a *non-trivial language* is one with at least three fluents.

Theorem 4.3.1 *(First Impossibility Theorem)*
Given a non-trivial language $\mathcal{B}$, there exists a causal system $\mathcal{Q}$ and (initial) state $w \in \mathcal{W}$ such that no ordering $<_w$ on states (generated by $\mathcal{B}$) satisfies (P′).

Proof: Assume that $\mathcal{B}$ has three propositional letters $\mathtt{a},\ \mathtt{b},\ \mathtt{c}$. Let the initial state be $w = \{\mathtt{a},\ \mathtt{b},\ \mathtt{c}\}$ and define s_1, s_2, s_3 and s_4 to be the following states: $s_1 = \{\neg\mathtt{a},\ \mathtt{b},\ \mathtt{c}\}$, $s_2 = \{\mathtt{a},\ \neg\mathtt{b},\ \mathtt{c}\}$, $s_3 = \{\neg\mathtt{a},\ \neg\mathtt{b},\ \mathtt{c}\}$ and $s_4 = \{\neg\mathtt{a},\ \mathtt{b},\ \neg\mathtt{c}\}$. Finally let $\mathcal{Q}$ be the following causal system: $\mathcal{Q} = \{\neg\mathtt{a} \wedge \mathtt{c} \Rightarrow \neg\mathtt{b}, \neg\mathtt{b} \wedge \mathtt{c} \Rightarrow \neg\mathtt{a}, \neg\mathtt{a} \wedge \mathtt{b} \Rightarrow \neg\mathtt{c}\}$.

Consider now the following direct effects (post-conditions) of actions. $\Delta_1 = \neg\mathtt{a} \wedge \mathtt{c}$, $\Delta_2 = \neg\mathtt{b} \wedge \mathtt{c}$, $\Delta_3 = (\mathtt{b} \leftrightarrow \neg\mathtt{c})$ and $\Delta_4 = (\bigwedge s_1) \vee (\bigwedge s_2) \vee (\bigwedge s_3) \vee (\bigwedge s_4)$. Clearly, states s_1 and s_3 satisfy Δ_1; s_2 and s_3 satisfy Δ_2; s_3 and s_4 satisfy Δ_3; and all four states s_1, s_2, s_3, s_4 satisfy Δ_4.

Suppose a (transitive) ordering on states $<_w$ satisfying condition (P′) exists. Now, the following is easily (albeit tediously) verified. $Res_{\mathcal{Q}}(w,\ \Delta_1) = \{s_3\}$ from which we conclude $s_1 \not<_w s_3$. $Res_{\mathcal{Q}}(w,\ \Delta_2) = \{s_3\}$, therefore $s_2 \not<_w s_3$. $Res_{\mathcal{Q}}(w,\ \Delta_3) = \{s_3,\ \{\mathtt{a},\ \mathtt{b},\ \neg\mathtt{c}\}\ \}$, therefore $s_4 \not<_w s_3$. Finally, $Res_{\mathcal{Q}}(w,\ \Delta_4) = \{s_4\}$ from which it follows that $(s_1 <_w s_3) \vee (s_2 <_w s_3) \vee (s_4 <_w s_3)$. This leads us to a contradiction. ■

The following impossibility result now follows quite straightforwardly and is more appropriate for our purposes given that condition (P) is a more faithful rendering of the spirit of preferential semantics than condition (P′). It allows us to conclude that a traditional preferential semantics (captured by condition (P)) cannot, in general, be given to McCain and Turner's causal theory of actions.

Theorem 4.3.2 *(Second Impossibility Theorem)*
Given a non-trivial language $\mathcal{B}$, there exists a causal system $\mathcal{Q}$ and (initial) state $w \in \mathcal{W}$ such that no ordering $<_w$ on states (generated by $\mathcal{B}$) satisfies (P).

We shall not give away preferential semantics entirely however. Our aim now becomes to retain as much of preferential semantics as possible and include a further mechanism to capture the influence of causality. To this end we investigate separate (though related) mechanisms for selecting successor states.

4.4 State-Selection Mechanisms

Taking a step backwards for a moment, we can simply view McCain and Turner's approach as a *state-selection mechanism*. More specifically, McCain and Turner's causal theory of actions, given some domain knowledge in terms of a causal theory $\mathcal{Q}$, specifies a way of selecting a subset $Res_{\mathcal{Q}}(w,\ E)$ of $[E]$ given an initial state w and direct effects E. $Res_{\mathcal{Q}}(w,\ E)$ returns exactly those states that are possible upon the occurrence of an action with direct effects E at state w.

Viewing this as a *selection function* however, we consider $Res_{\mathcal{Q}}(w,\ E)$ to be a function selecting the 'best' E-states (with respect to w). This is the view we shall adopt here in presenting two further state-selection mechanisms: *state elimination systems* and *state transition systems*. State transition systems will provide a variant of the augmented preferential semantics we seek in terms of our aims.

This desired result is achieved in two steps.We begin by showing how to inter-translate McCain and Turner causal systems and state elimination systems in a way that preserves the selection process. We then show how to inter-translate state elimination systems and state transition systems (again preserving the selection process). In truth, we could do away with state elimination systems and simply translate directly between causal systems and state transition systems. However, we choose not to do so because it simplifies the proofs and provides further insight into the nature of context-sensitive causality captured by McCain and Turner's approach.

4.5 State Elimination Systems

In this section we describe our first state-selection mechanism: state elimination systems. The underlying idea is to use *state elimination rules* to discard E-states from further consideration for we have noted above that in McCain and Turner's [37] causal theory

$$Res_{\mathcal{Q}}(w,\ E) \subseteq [E].$$

A state rejected or eliminated by a state elimination rule is one which contravenes a causal rule deemed to hold in the successor state (in fact, in the causal system as a whole).

Definition 4.5.1 *(State elimination rule)*
A state elimination rule *(or simply,* elimination rule*) is an expression of the form* $\{r_1, r_2, \cdots, r_k, r_{k+1}, \cdots, r_n\} \rhd \{r_1, r_2, \cdots, r_k\}$ *where each* r_i *is a state.*

A state elimination system $\mathcal{S}$ is a set of state elimination rules. An elimination rule functions by rejecting certain states from among those currently considered possible. Suppose that according to an agent's current beliefs it considers the states that are possible to be among $\{r_1, \cdots, r_n\}$. An elimination rule like that in Definition 4.5.1 allows the agent to reject states $r_{k+1}, \cdots, r_n$.

Let us briefly consider the mechanics of a state elimination system. At any point we are working with the set of states currently being entertained (a subset of $[E]$). We repeatedly apply elimination rules to this set of states to reject the illegitimate ones focussing on the possible successor states. All elimination rules need to be applied until

no further states can be rejected to ensure that all illegitimate states have been purged and only definite possibilities remain. To put it another way, a state elimination system acts as a *filtering* mechanism; illegitimate states are successively filtered out through use of elimination rules.

Definition 4.5.2 *($\rightsquigarrow$ and $\overset{*}{\rightsquigarrow}$)*
In a state elimination system $\mathcal{S}$, we shall say that a set of states Q yields *a set of states R* in one step, *denoted by $Q \rightsquigarrow R$, if and only if there exists an elimination rule $X \triangleright Y$ such that $Q \subseteq X$ and $R = Q \cap Y$. We define $\overset{*}{\rightsquigarrow}$ to be the reflexive transitive closure of $\rightsquigarrow$.*

After the application of certain elimination rules we find that any further application does not result in the rejection of additional states. At this point we reach a *final set of states*; a point of equilibrium.

Definition 4.5.3 *(Final state)*
A set of states Q is final *in $\mathcal{S}$ if and only if for any R such that $Q \overset{*}{\rightsquigarrow} R$, it follows that $Q = R$. If Q is a singleton and final, we will call the state in Q final.*

One last notion that we require is that of an *E-predecessor* of a given state; those E-states preceding the given state with respect to an ordering based on symmetric difference. More formally:

Definition 4.5.4 *(E-predecessor)*
Given any two states $w,\ r$ and any sentence E, the E-predecessors of r with respect to w is defined to be the set

$$\langle\!\langle r, E \rangle\!\rangle_w = \{r' :\ r' \in [E] \text{ and } \mathit{Diff}(w, r') \subseteq \mathit{Diff}(w, r)\},$$

where $\mathit{Diff}(x,\ y)$ denotes the symmetric difference of states x and y.

In other words, the E-predecessors are those E-states that are at least as close to w as r:

$$\langle\!\langle r, E \rangle\!\rangle_w = \{r' :\ r' \in [E] \text{ and } r' \prec_w r\},$$

where $\prec_w$ is the PMA ordering with respect to the state w — in other words, $r \prec_w s$ if and only if $\mathit{Diff}(w, r) \subseteq \mathit{Diff}(w, s)$ [70].

It is clear that any $r \in [E]$ is an E-predecessor of itself with respect to w (i.e., $r \in \langle\!\langle r, E \rangle\!\rangle_w$). The E-predecessors of r with respect to w are just the E-states which agree with w on at least those fluents where w and r agree and possibly others. Let us consider the set $min(\prec_w, [E])$ defined as a subset of $[E]$ containing states nearest to the set w in terms of the ordering $\prec_w$:

$$min(\prec_w, [E]) = \{p \in [E], \neg\exists\ q \in [E],\ q \neq p,\ q \prec_w p\}.$$

One may compare this set with the set $\langle\!\langle r, E \rangle\!\rangle_w$. The latter set includes all E-predecessors of a particular state r, while the set $min(\prec_w, [E])$ contains *all* $\prec_w$-minimal states in $[E]$, some of which may not be E-predecessors of state r.

We are now in a position to define a state-selection mechanism based on state elimination systems.

Definition 4.5.5 ($\texttt{Next}_{\mathcal{S}}(w,\ E)$)
With any state elimination system $\mathcal{S}$ we associate a result function $\texttt{Next}_{\mathcal{S}}$, mapping a final in $\mathcal{S}$ state w and a sentence E to the set of states $\texttt{Next}_{\mathcal{S}}(w,\ E)$, and defined as follows:

$$\texttt{Next}_{\mathcal{S}}(w,\ E) = \{r \in [E] :\ r \text{ is final in } \mathcal{S} \text{ and } \langle\!\langle r, E \rangle\!\rangle_w \overset{*}{\rightsquigarrow} \{r\}\ \}.$$

In the following section we characterise $Res_{\mathcal{Q}}(w,\ E)$ in terms of $\texttt{Next}_{\mathcal{S}}(w,\ E)$. First, however, let us briefly consider the definition of $\texttt{Next}_{\mathcal{S}}(w,\ E)$. According to the definition above, a state r is a possible resultant state if and only if all its E-predecessors (with respect to w) are rejected by elimination rules in $\mathcal{S}$ but r and only r is retained. If r is retained along with some other state, then there is some closer state (one with 'less' change) consistent with the state elimination system $\mathcal{S}$ (and, therefore, causal system $\mathcal{Q}$) under consideration. Moreover, it means that there is something in the state(s) for which causality cannot account. If r is rejected on the other hand, it must violate a causal rule. For these reasons we only consider the E-predecessors of r to determine whether it belongs to $\texttt{Next}_{\mathcal{S}}(w,\ E)$; we need to determine whether r is illegitimate or whether a 'closer' state satisfies the causal rules. If either is the case, we can safely reject the state. Otherwise, we can retain the state.

4.5.1 Causal Systems and State Elimination Systems

We now establish the interrelationship between causal systems and state elimination systems. This will give us a way of moving back and forth between the two systems facilitating the final inter-translation between causal systems and state transition systems.

First, however, we make the following helpful observation.

Lemma 4.5.6 *For any two states r, w and sentence E,*

$$[(w \cap r) \cup \{E\}] = \langle\!\langle r, E \rangle\!\rangle_w.$$

We now turn to the main result of this section. A state elimination system can exactly capture a causal system (and vice versa). In other words, selection-equivalence can be achieved between these two kinds of selection functions.

Theorem 4.5.7 *For every causal system there exists a selection-equivalent state elimination system. Conversely, for every state elimination system there exists a selection-equivalent causal system.*

Proof: (Sketch[4])

($\Rightarrow$) Let $\mathcal{Q}$ be an arbitrary causal system. For every causal law $\varphi \Rightarrow \psi$ in $\mathcal{Q}$, produce the elimination rule $[\varphi] \rhd [\varphi \wedge \psi]$. Call $\mathcal{S}$ the set of elimination rules so produced. Using lemma 4.5.6, we can verify that for any legitimate state w and sentence E, $Res_{\mathcal{Q}}(w,\ E) = \mathtt{Next}_{\mathcal{S}}(w,\ E)$.

($\Leftarrow$) Let $\mathcal{S}$ be an arbitrary state elimination system. For every elimination rule $X \rhd Y$ produce the causal law $\varphi \Rightarrow \psi$, where $\varphi,\ \psi$ are such that $[\varphi] = X$ and $[\psi] = Y$ (since our language is a finitary propositional one, such φ and ψ always exist). The set of causal laws so produced, call it $\mathcal{Q}$, is selection-equivalent to $\mathcal{S}$.

■

Of particular note is the relationship between causal rules and elimination rules: $\phi \Rightarrow \psi$ if and only if $[\phi] \rhd [\phi \wedge \psi]$, or, equivalently, $[\phi] \rhd [\phi]\ \cap\ [\psi]$.

We can also identify an important class of state elimination systems that will be useful later.

[4] As usual, a detailed proof is given in the Appendix.

Definition 4.5.8 *($\mathcal{S}$ Unary)*
A state elimination system $\mathcal{S}$ is unary *if and only if every elimination rule eliminates precisely one state, i.e. for all $(X \triangleright Y) \in \mathcal{S}$, $X \setminus Y$ is a singleton.*

The following result reveals an interesting and important aspect of unary state elimination systems.

Theorem 4.5.9 *Every state elimination system is selection-equivalent to a unary state elimination system.*

4.6 State Transition Systems

In this section we consider our second state-selection mechanism: state transition systems. Again, we shall obtain a direct characterisation of causal systems (and state elimination systems). In this case we have a preferential mechanism augmented by further structure to achieve the desired result.

A state transition system consists of a binary relation on states intended to represent possible transitions between states due to the presence of causality. It is this relation that, together with a preferential structure based on symmetric difference (PMA ordering), will be used to determine successor states. In other words, a state transition system can be described with a simple variant of the *augmented preferential semantics* (using the approximation $\mathcal{W} = \Gamma$, in particular).

We begin with some requisite definitions.

Definition 4.6.1 *A* state transition system *$\mathcal{M}$ is an irreflexive binary relation on the set $\mathcal{W}$ of states (i.e., $\mathcal{M} \subseteq \mathcal{W} \times \mathcal{W}$). We shall say that a state r is* final *in $\mathcal{M}$ if and only if there is no state r' such that $\mathcal{M}(r, r')$.*

As argued in Chapter 2, the binary relation $\mathcal{M}$ can be considered to represent state transitions due to the influence of causality.

We are now in a position to define the mechanism for selecting successor states $\texttt{Succ}_{\mathcal{M}}(w,\ E)$ for a state transition system $\mathcal{M}$ given initial state w and an action with direct effects E. It will be shown later that the selection function $\texttt{Succ}_{\mathcal{M}}(w,\ E)$ can

be characterised according to the augmented preferential semantics. At this stage, as mentioned before, we treat this function merely as a selection mechanism.

Definition 4.6.2 ($\texttt{Succ}_{\mathcal{M}}(w,\ E)$)
To any state transition system $\mathcal{M}$ we associate a function $\texttt{Succ}_{\mathcal{M}}$, mapping a final (in $\mathcal{M}$) state w and a sentence E to the set of states $\texttt{Succ}_{\mathcal{M}}(w,\ E)$, defined as follows:

$$\texttt{Succ}_{\mathcal{M}}(w,\ E) = \{r \in [E] :\ r \text{ is final in } \mathcal{M} \text{ and}$$
$$\text{there is a Hamiltonian path through all states in } \langle\!|r, E|\!\rangle_w\}.$$

A Hamiltonian path is one which traverses every vertex (here a state) of a graph [71]. In this case, the graph's vertices are the E-predecessors of r and the edges are given by the binary relation $\mathcal{M}$ — i.e., there is an edge between states p and q if and only if $\mathcal{M}(p, q)$. The significance of a Hamiltonian path will be considered further in the next section.

It is important to notice that $\texttt{Succ}_{\mathcal{M}}(w,\ E)$ is determined by two components: a preference ordering on states, based on symmetric difference, used to derive the E-predecessors of r with respect to w (i.e., $\langle\!|r, E|\!\rangle_w$) and the binary relation on states $\mathcal{M}$. We maintain that the preference ordering captures the Principle of Minimal Change while the binary relation captures the effect of causality. Notice firstly that a state must be reachable via a Hamiltonian path through all E-predecessors ending in r. If this is not possible, either a 'closer' E-state is consistent with the causal rules that hold (absence of Hamiltonian path) or r violates a causal rule (path does not end at r; i.e., r is not final). Another important point is that, like state elimination systems, we only need to consider E-predecessors of r (with respect to w) in order to determine whether it is a successor state.

4.6.1 State Elimination Systems and State Transition Systems

In this section we establish an inter-translation between state elimination systems and state transition systems. We can then use the results of Section 4.5.1 to establish a correspondence between causal systems and state transition systems. This gives us the result we seek: an augmented preferential semantics for McCain and Turner's causal theory of actions.

First we require some definitions.

Definition 4.6.3 *(Dissolvable set)*
We shall say that the string of elimination rules $\sigma_1; \sigma_2; \cdots; \sigma_n$ from a unary system $\mathcal{S}'$ dissolves *an arbitrary set of states Π with cardinality $n+1$ if and only if after applying these rules successively (in the order given), all but one of the states of Π are eliminated and, furthermore, the one remaining state is final in $\mathcal{S}'$. We shall also call Π a* dissolvable *set of states.*

Essentially, a Hamiltonian path goes through states as they are eliminated by unary elimination rules.

Definition 4.6.4 *(Trace)*
We shall call the sequence of states $r_1; r_2; \cdots; r_n; w$ a trace *for a set Π (in $\mathcal{S}'$) if and only if the string $\sigma_1; \sigma_2; \cdots; \sigma_n$ dissolves Π, and for all $1 \leq i \leq n$, state r_i is the state of Π that is eliminated by the rule σ_i, while state w is the only state of Π that is not eliminated.*

Now we are ready to formulate the following characterisation.

Theorem 4.6.5 *For every state elimination system $\mathcal{S}$ there is a selection-equivalent state transition system $\mathcal{M}$. Conversely, for every state transition system $\mathcal{M}$ there is a selection-equivalent state elimination system $\mathcal{S}$.*

Proof: (Sketch)
($\Longrightarrow$) Let $\mathcal{S}$ be a state elimination system. Let $\mathcal{S}'$ be a unary state elimination system that is selection-equivalent to $\mathcal{S}$. From $\mathcal{S}'$ we construct a selection-equivalent state transition system $\mathcal{M}$ in the following manner.

For any two states r and r', we shall specify $\mathcal{M}(r, r')$ if and only if there is a dissolvable set of states Π containing r and r', such that for some trace of Π in $\mathcal{S}'$, r' appears immediately after r.

It can be shown that $\mathcal{M}$ is selection-equivalent to $\mathcal{S}'$.
($\Longleftarrow$) Proved by reversing the construction presented above.

■

Note also that the construction used in the proof ensures that the binary relation $\mathcal{M}$ is irreflexive. Irreflexivity may not seem obvious because, given a simple elimination rule $\{r_1, r_2\} \rhd \{r_1\}$, we derive (by Definition 4.5.2) that $\{r_1, r_2\} \rightsquigarrow \{r_1\}$ and $\{r_1\} \rightsquigarrow \{r_1\}$ (as well as $\{r_2\} \rightsquigarrow \emptyset$). However, any trace is a result of elimination of n states by n elimination rules dissolving a set of cardinality $n + 1$ (by Definition 4.6.4). Since the procedure is applied via a unary elimination system, each rule eliminates precisely one state. In other words, we cannot accept a possibility that one rule eliminates two states or more, while some other rule does not eliminate any state at all. Therefore, we conclude that any given trace can not contain a state r followed by itself, ensuring irreflexivity of $\mathcal{M}$ despite the fact that some states may be reflexive with respect to $\rightsquigarrow$.

The central result of this chapter, as expressed by the following corollary, is now obtained by combining theorems 4.5.7 and 4.6.5.

Corollary 4.6.6 *For every causal system $\mathcal{Q}$ there exists a selection-equivalent state transition system $\mathcal{M}$. Conversely, for every state transition system $\mathcal{M}$ there exists a selection-equivalent causal system $\mathcal{Q}$.*

This result states that it is possible to exactly characterise McCain and Turner's causal theory of actions with a variant of the preferential-style semantics (where the preferential structure is defined in terms of symmetric difference) augmented with a binary relation on states and through the notion of a Hamiltonian path. The precise characterisation will be described in Chapter 7, where a particular gradient is specified in order to capture predecessor states $\langle\!\langle r, E \rangle\!\rangle_w$.

4.6.2 Detailed Examples

We begin with a basic example, involving a very simple single constraint (causal rule), and consider a couple of actions, the second one being more complex than the first.

Example 4.6.7 *Consider a domain with three fluents a, b, c, and eight states: $\{a, b, c\}$, $\{a, b, \neg c\}$, $\{a, \neg b, c\}$, $\{\neg a, b, c\}$, $\{a, \neg b, \neg c\}$, $\{\neg a, b, \neg c\}$, $\{\neg a, \neg b, c\}$ and $\{\neg a, \neg b, \neg c\}$. Consider also the following causal rule*

$$\neg b \;\Rightarrow\; \neg c.$$

The specified causal rule excludes states $\{a, \neg b, c\}$ and $\{\neg a, \neg b, c\}$ from the set of legitimate states $\mathcal{D}$, and produces the following elimination rule:

$$[\neg b] \rhd [\neg b \wedge \neg c]$$

or

$$\{\{a, \neg b, c\}, \{\neg a, \neg b, c\}, \{a, \neg b, \neg c\}, \{\neg a, \neg b, \neg c\}\} \rhd \{\{a, \neg b, \neg c\}, \{\neg a, \neg b, \neg c\}\}.$$

This elimination rule, in turn, is (selection-)equivalent to the following unary elimination rules:

$$\{\{\neg a, \neg b, c\}, \{a, \neg b, \neg c\}, \{\neg a, \neg b, \neg c\}\} \rhd \{\{a, \neg b, \neg c\}, \{\neg a, \neg b, \neg c\}\},$$

$$\{\{a, \neg b, c\}, \{a, \neg b, \neg c\}, \{\neg a, \neg b, \neg c\}\} \rhd \{\{a, \neg b, \neg c\}, \{\neg a, \neg b, \neg c\}\},$$

$$\{\{a, \neg b, c\}, \{\neg a, \neg b, c\}, \{\neg a, \neg b, \neg c\}\} \rhd \{\{\neg a, \neg b, c\}, \{\neg a, \neg b, \neg c\}\},$$

$$\{\{a, \neg b, c\}, \{\neg a, \neg b, c\}, \{\neg a, \neg b, \neg c\}\} \rhd \{\{a, \neg b, c\}, \{\neg a, \neg b, \neg c\}\},$$

$$\{\{a, \neg b, c\}, \{\neg a, \neg b, c\}, \{a, \neg b, \neg c\}\} \rhd \{\{a, \neg b, c\}, \{a, \neg b, \neg c\}\},$$

$$\{\{a, \neg b, c\}, \{\neg a, \neg b, c\}, \{a, \neg b, \neg c\}\} \rhd \{\{\neg a, \neg b, c\}, \{a, \neg b, \neg c\}\},$$

$$\{\{\neg a, \neg b, c\}, \{a, \neg b, \neg c\}\} \rhd \{\{a, \neg b, \neg c\}\},$$

$$\{\{\neg a, \neg b, c\}, \{\neg a, \neg b, \neg c\}\} \rhd \{\{\neg a, \neg b, \neg c\}\},$$

$$\{\{a, \neg b, c\}, \{a, \neg b, \neg c\}\} \rhd \{\{a, \neg b, \neg c\}\},$$

$$\{\{a, \neg b, c\}, \{\neg a, \neg b, \neg c\}\} \rhd \{\{\neg a, \neg b, \neg c\}\},$$

$$\{\{\neg a, \neg b, c\}\} \rhd \emptyset,$$

$$\{\{a, \neg b, c\}\} \rhd \emptyset.$$

These unary elimination rules produce the following traces

$$\{a, \neg b, c\}; \{\neg a, \neg b, c\}; \{a, \neg b, \neg c\},$$

$$\{a, \neg b, c\}; \{\neg a, \neg b, c\}; \{\neg a, \neg b, \neg c\},$$

$$\{\neg a, \neg b, c\}; \{a, \neg b, c\}; \{a, \neg b, \neg c\},$$

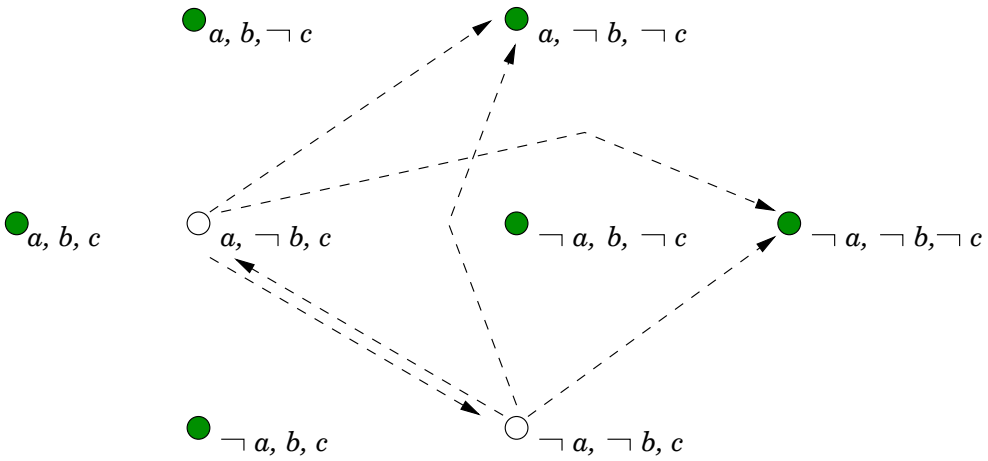

Figure 4.1: Binary relation $\mathcal{M}$ constructed from the causal rule $\neg b \Rightarrow \neg c$.

$$\{\neg a, \neg b, c\}; \{a, \neg b, c\}; \{\neg a, \neg b, \neg c\}.$$

where the states $\{a, \neg b, \neg c\}$ and $\{\neg a, \neg b, \neg c\}$ are final. Therefore, one may construct the binary relation $\mathcal{M}$ (Figure 4.1), including

$$\mathcal{M}(\{a, \neg b, c\}, \{\neg a, \neg b, c\}),$$

$$\mathcal{M}(\{\neg a, \neg b, c\}, \{a, \neg b, c\}),$$

$$\mathcal{M}(\{\neg a, \neg b, c\}, \{a, \neg b, \neg c\}),$$

$$\mathcal{M}(\{\neg a, \neg b, c\}, \{\neg a, \neg b, \neg c\}),$$

$$\mathcal{M}(\{a, \neg b, c\}, \{a, \neg b, \neg c\}),$$

$$\mathcal{M}(\{a, \neg b, c\}, \{\neg a, \neg b, \neg c\}).$$

Simple Action

Let us consider an action with the post-condition $\neg b$, executed in the state $w = \{a, b, c\}$. According to the causal fixed-points approach,

$$Res_{\mathcal{Q}}(w, \neg b) = \{\{a, \neg b, \neg c\}\}.$$

Note that the state $r = \{\neg a, \neg b, \neg c\}$ is not a fixed-point because $(w \cap r) \cup \{\neg b\} = \{\neg b\}$ is not sufficient to explain all literals in r.

It is easy to verify that

$$\texttt{Succ}_{\mathcal{M}}(w,\ \neg b) = \{\{a, \neg b, \neg c\}\}.$$

In other words, the selection function of the state transition systems selects the same successor state $p = \{a, \neg b, \neg c\}$ — there is a Hamiltonian path through $\neg b$-predecessors of p:

$$\mathcal{M}(\{a, \neg b, c\}, \{a, \neg b, \neg c\}).$$

The state r, on the other hand, is not selected because it is not $\mathcal{M}$-reachable from p, which is its PMA predecessor: $p \prec_w r$.

Complex Action

Now let us consider an action with the more complex post-condition $E = (\neg b \wedge c) \vee (\neg a \wedge \neg b \wedge \neg c)$, executed in the same initial state $w = \{a, b, c\}$. According to the causal fixed-points approach,

$$Res_{\mathcal{Q}}(w,\ E) = \{\{\neg a, \neg b, \neg c\}\}.$$

Note that the state $r = \{\neg a, \neg b, \neg c\}$ is a fixed-point because $(w \cap r) \cup E = \{(\neg b \wedge c) \vee (\neg a \wedge \neg b \wedge \neg c)\}$ entails $\neg b$, which in turn causally infers $\neg c$, and $\neg b \wedge \neg c$ together with E entail $\neg a$ as well (so that all literals in r are explained).

It is easy to verify that

$$\texttt{Succ}_{\mathcal{M}}(w,\ E) = \{\{\neg a, \neg b, \neg c\}\}.$$

In other words, the selection function of the state transition systems selects the same successor state r. This happens to be the case because there is a Hamiltonian path through the E-predecessors of r, ending in r — more precisely,

$$\mathcal{M}(\{a, \neg b, c\}, \{\neg a, \neg b, c\}),$$
$$\mathcal{M}(\{\neg a, \neg b, c\}, \{\neg a, \neg b, \neg c\}).$$

Action with Multiple Successor States

Let us consider an action with the post-condition $E' = a \leftrightarrow \neg b$, executed in the state $w = \{a, b, c\}$. According to the causal fixed-points approach,

$$Res_{\mathcal{Q}}(w,\ E') = \{\{\neg a, b, c\}, \{a, \neg b, \neg c\}\}.$$

Note that two successor states are not comparable in terms of the PMA.

It is easy to verify that

$$\mathtt{Succ}_{\mathcal{M}}(w,\ E') = \{\{\neg a, b, c\}, \{a, \neg b, \neg c\}\}.$$

In other words, the selection function of the state transition systems selects the same successor states. On the one hand, there is a Hamiltonian path through E'-predecessors of the state $\{a, \neg b, \neg c\}$:

$$\mathcal{M}(\{a, \neg b, c\}, \{a, \neg b, \neg c\}).$$

The state $y = \{\neg a, b, c\}$, on the other hand, is selected because it is the only state in $\langle\!\langle y, E' \rangle\!\rangle_w$, being in addition a trivially stable state.

The following example is described by Zhang [72], and is aimed to illustrate "mutual effects occurring between causal relations and logical constraints". In other words, we trace causal propagation unfolding in the presence of logical domain constraints.

Example 4.6.8 *Consider a domain with five fluents* on_1, on_2, $light$, $bright(room)$, $visible(painting)$, *the following causal rules*

$$on_1 \wedge on_2 \ \Rightarrow\ light,$$

$$bright(room) \ \Rightarrow\ visible(painting),$$

and a single logical constraint represented without loss of generality as

$$True \ \Rightarrow\ light \supset bright(room).$$

The first rule states that whenever two switches are on, the lamp will light, while the second rule says that if the room is bright, then the painting on the wall is visible. The last domain constraint states that the fact that the light is on implies the fact that the room is bright (of course, the contrapositive formula $\neg bright(room) \supset \neg light$ is implied as well, unlike the contrapositives of the causal rules).

The initial state is

$$w = \{\neg bright(room), \neg light, \neg on_1, on_2, \neg visible(painting)\}.$$

Consider the action of toggling the first switch with the post-condition on_1. The following two states are of particular interest:

$$s_1 = \{bright(room), light, on_1, on_2, visible(painting)\}$$

$$s_2 = \{\neg bright(room), \neg light, on_1, \neg on_2, \neg visible(painting)\}$$

One of the reasons for considering s_1 and s_2 (besides the facts that both states satisfy the post-condition and are legitimate) is their selection by a particular policy of causal minimisation — the method based on minimisation of causal change sets, proposed by Brewka and Hertzberg [4]. While the state s_1 is a desired successor, the state s_2 is counter-intuitive. A description of Brewka and Hertzberg's approach is outside the scope of this work. We just point out that the critique of this policy, presented by Zhang [72], highlighted its shortcomings mainly as follows: the definition of causal change set does not properly handle causal inference "disconnected" by a logical implication.

According to the causal fixed-points approach,

$$Res_{\mathcal{Q}}(w,\ on_1) = \{s_1\}.$$

It is easy to verify that $(w \cap s_1) \cup \{on_1\} = \{on_1, on_2\}$ causally infers $light$, which in turn implies $bright(room)$, followed by causal inference of $visible(painting)$.

The state s_2 is not a fixed-point because

$$(w \cap s_2) \cup \{on_1\} = \{\neg bright(room), \neg light, on_1, \neg visible(painting)\}$$

is not sufficient to infer the literal $\neg on_2$ present in s_2.

Our intention, at this stage, is to demonstrate that

$$\mathtt{Succ}_{\mathcal{M}}(w,\ on_1) = \{s_1\}.$$

The elimination rules corresponding to causal rules are as follows:

$$[on_1 \wedge on_2] \rhd [on_1 \wedge on_2 \wedge light]$$

$$[bright(room)] \rhd [bright(room) \wedge visible(painting)]$$

$$\mathcal{W} \rhd [light \supset bright(room)],$$

where $\mathcal{W}$ is the set of all 32 states.

We will need one particular trace ending in s_1, produced by these rules:

$$\{\neg bright(room), \neg light, on_1, on_2, \neg visible(painting)\};$$

$$\{\neg bright(room), \neg light, on_1, on_2, visible(painting)\};$$

$$\{bright(room), \neg light, on_1, on_2, \neg visible(painting)\};$$

$$\{bright(room), \neg light, on_1, on_2, visible(painting)\};$$

$$\{bright(room), light, on_1, on_2, \neg visible(painting)\};$$

$$\{\neg bright(room), light, on_1, on_2, \neg visible(painting)\};$$

$$\{\neg bright(room), light, on_1, on_2, visible(painting)\};$$

$$\{bright(room), light, on_1, on_2, visible(painting)\}.$$

To check that this is indeed a trace, note that

- the first four states are eliminated by the first elimination rule (more precisely, corresponding unary rules), because they all entail $on_1 \wedge on_2$ but not $light$;
- the fifth state is eliminated by the second elimination rule (or its unary derivations), because it contains $bright(room)$ but not $visible(painting)$;
- the sixth and the seventh states are eliminated by the last elimination rule (or its unary "offspring"), because they do not satisfy $light \supset bright(room)$.

Of course, other traces are possible among these eight states, but the state s_1 is invariably a final state (with respect to the generated relation $\mathcal{M}$). To show that the state s_1 is a successor state for the state w and the toggle action on_1, we note that all its predecessors $\langle\!\langle s_1, on_1 \rangle\!\rangle_w$ are precisely the states in the analysed trace. Therefore, there exists a Hamiltonian path through the states in $\langle\!\langle s_1, on_1 \rangle\!\rangle_w$, ending in s_1.

We needed to show that

$$\mathtt{Succ}_{\mathcal{M}}(w,\ on_1) = \{s_1\}.$$

The selection of only s_1 out of all the states consistent with $on_1 \wedge on_2$ means that now we need to verify that none of the states consistent with $on_1 \wedge \neg on_2$ is selected by $\mathtt{Succ}_{\mathcal{M}}(w,\ on_1)$ (exhausting all states consistent with the post-condition on_1). In particular, we need to verify that

$$s_2 = \{\neg bright(room), \neg light, on_1, \neg on_2, \neg visible(painting)\}$$

is not a successor state.

The set of $[on_1]$-consistent predecessors of s_2 is $\langle\!\langle s_2, on_1 \rangle\!\rangle_w = \{s_3, s_2\}$, where

$$s_3 = \{\neg bright(room), \neg light, on_1, on_2, \neg visible(painting)\}.$$

This set cannot be dissolved to s_2, because there are no elimination rules at all where s_2 happens to be on the right-hand side. Hence, there is no Hamiltonian path in $\langle\!\langle s_2, on_1 \rangle\!\rangle_w$. Analogously, it can be easily shown that for any other state s consistent with $on_1 \wedge \neg on_2$ there is no Hamiltonian path in $\langle\!\langle s, on_1 \rangle\!\rangle_w$.

Therefore, s_1 is the only successor state selected by $\mathtt{Succ}_{\mathcal{M}}(w,\ on_1)$.

Interestingly, if the constraint

$$True \ \Rightarrow\ light \supset bright(room)$$

is removed, and another action with the post-condition $E = on_1 \wedge (light \supset bright(room))$ is considered, we obtain the same results:

$$Res_{\mathcal{Q}}(w,\ E) = \mathtt{Succ}_{\mathcal{M}}(w,\ E) = \{s_1\}.$$

In other words, the selection functions $Res_{\mathcal{Q}}(w,\ E)$ and $\mathtt{Succ}_{\mathcal{M}}(w,\ E)$ are syntax-independent.

4.7 Discussion and Outlook

In this chapter we attempted to determine whether it is possible to provide McCain and Turner's [37] causal theory of actions with a preferential semantics in the spirit of Shoham [60]. We demonstrated, through use of an impossibility theorem (Theorem 4.3.2), that this is not possible in general when we do not extend the original language $\mathcal{B}$ and assume that the preferential ordering satisfies transitivity. Choosing not to abandon preferential semantics entirely, we then introduced two state-selection mechanisms: state elimination systems and state transition systems. The latter of these provides the target semantics augmenting a preferential structure based on symmetric difference with a binary relation on states, and making use of the notion of a Hamiltonian path. The former provides further insight into the nature of context-sensitive causality used in McCain and Turner's approach. Importantly, we show that causal systems, state elimination systems and state transition systems, as defined here, are all (selection-) equivalent.

In summary, it is possible to retain preferential semantics and augment it in capturing the causal theory of McCain and Turner. We maintain that the preferential component of state transition systems relates to the Principle of Minimal Change while the binary relation on states relates to the Principle of Causal change. It is our contention that both of these components — minimal change and causality — are required if one is to supply a *concise* solution to the Frame and Ramification problems; the two components can co-exist and, in fact, complement each other.

The notion of a Hamiltonian path through E-predecessors in a state transition system is an interesting one. Essentially, a Hamiltonian path serves as a contextual mechanism much in the same way that augmenting the underlying language through the addition of extra predicates does. The additional information allows the domain causality to contribute in certain situations and not in others.

It is interesting to compare the selection function $\texttt{Succ}_{\mathcal{M}}(w,\ E)$ of state transition systems with another selection function that does not use a Hamiltonian path through predecessor states. Instead, this function, denoted $\texttt{Reach}_{\mathcal{M}}(w,\ E)$, simply requires that a successor state is $\mathcal{M}$-reachable from every predecessor in $\langle\!\langle r, E \rangle\!\rangle_w$. More precisely, we define it as follows.

Definition 4.7.1 (**$\texttt{Reach}_{\mathcal{M}}(w,\ E)$**)
To any state transition system $\mathcal{M}$ we associate a function $\texttt{Reach}_{\mathcal{M}}$, mapping a final in $\mathcal{M}$ state w and a sentence E to the set of states $\texttt{Reach}_{\mathcal{M}}(w,\ E)$, defined as follows:

$$\texttt{Reach}_{\mathcal{M}}(w,\ E) = \{r \in [E] :\ r \text{ is final in } \mathcal{M} \text{ and } \mathcal{M}^*(r', r) \text{ for all } r' \in \langle\!\langle r, E \rangle\!\rangle_w\}.$$

The function $\texttt{Reach}_{\mathcal{M}}(w,\ E)$ is intuitively appealing because it does not impose any additional topological requirements (like a Hamiltonian path) on the process of causal propagation. Besides, this function is quite similar to the selection function used in the simple variant of our augmented preferential semantics presented in Chapter 2. However, the function $\texttt{Reach}_{\mathcal{M}}(w,\ E)$ does not capture all causal fixed-points. It is easy to establish that by the definitions 4.6.2 and 4.7.1,

$$\texttt{Succ}_{\mathcal{M}}(w,\ E) \subseteq \texttt{Reach}_{\mathcal{M}}(w,\ E).$$

If there exists a Hamiltonian path through predecessor states ending in a state r then, obviously, the state r is transitively reachable from each predecessor individually. The reverse statement does not hold.

Moreover, state transition systems based on the function $\texttt{Reach}_{\mathcal{M}}(w,\ E)$ would not completely characterise the causal systems with fixed-points $Res_{\mathcal{Q}}(w,\ E)$, as shown by the following example.

4.7.1 An Example of Context-sensitivity

Consider a very simple domain with three fluents a, b, c, and eight states, out of which we label the following five: $w = \{a, b, c\}$, $s = \{a, b, \neg c\}$, $p = \{a, \neg b, \neg c\}$, $q = \{\neg a, b, \neg c\}$ and $r = \{\neg a, \neg b, \neg c\}$. Consider also the following causal rules

$$a \wedge \neg c \ \Rightarrow\ a \wedge \neg b \wedge \neg c,$$

$$b \wedge \neg c \ \Rightarrow\ \neg a \wedge b \wedge \neg c,$$

$$\neg b \wedge \neg c \ \Rightarrow\ \neg a \wedge \neg b \wedge \neg c,$$

$$\neg a \wedge \neg c \ \Rightarrow\ \neg a \wedge \neg b \wedge \neg c.$$

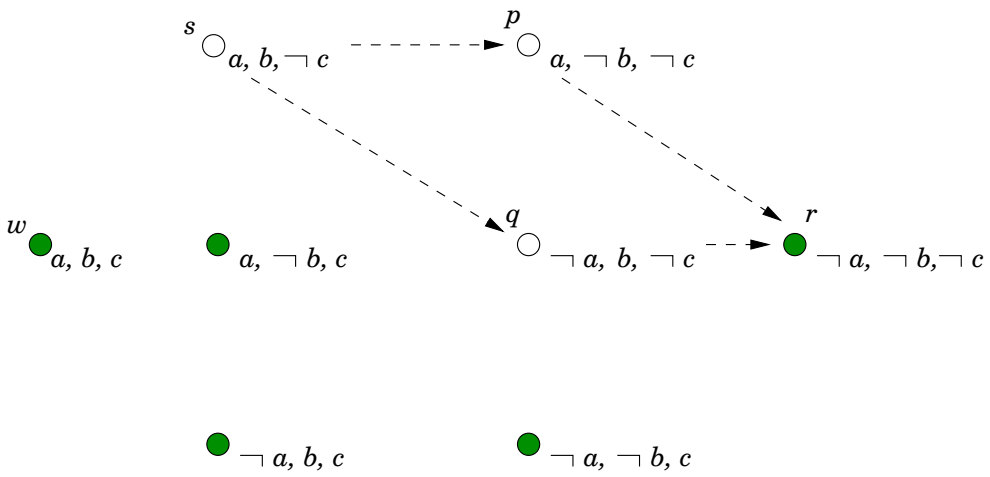

Figure 4.2: Absence of a Hamiltonian path.

These rules exclude states s, p and q from the legitimate states $\mathcal{D}$, and produce the following elimination rules:

$$\{s, p\} \rhd \{p\},$$

$$\{s, q\} \rhd \{q\},$$

$$\{p, r\} \rhd \{r\},$$

$$\{q, r\} \rhd \{r\}.$$

This results in two traces $s; p; r$ and $s; q; r$, where the state r is final. Therefore, one may construct the binary relation $\mathcal{M}$, including $\mathcal{M}(s, p)$, $\mathcal{M}(s, q)$, $\mathcal{M}(p, r)$ and $\mathcal{M}(q, r)$ (Figure 4.2).

Now, let us consider an action with the post-condition $\neg c$, executed in the state w. According to the causal fixed-points approach, there are no successor states. In other words,

$$Res_{\mathcal{Q}}(w,\ \neg c) = \emptyset.$$

Note that the state r is not a fixed-point because $(w \cap r) \cup \{\neg c\} = \{\neg c\}$ is not sufficient to explain all literals in r.

There are no successor states according to $\texttt{Succ}_{\mathcal{M}}(w,\ \neg c)$ as well — as there is no Hamiltonian path through the predecessors of the final state r in the set $\langle\!\langle r, \neg c \rangle\!\rangle_w$ =

$\{s, p, q, r\}$:

$$\texttt{Succ}_{\mathcal{M}}(w,\ \neg c) = \emptyset.$$

This is so despite the fact that r is $\mathcal{M}$-reachable from every predecessor in $\langle\!\langle r, \neg c \rangle\!\rangle_w$. In other words,

$$\texttt{Reach}_{\mathcal{M}}(w,\ \neg c) = \{r\}.$$

This clearly demonstrates that in order for a final state to be a causal fixed-point, it is not sufficient to be $\mathcal{M}$-reachable from every predecessor — there must be a Hamiltonian path through all the predecessors.

4.7.2 Causal Systems Closed under Disjunction

In this section, we sketch a certain sub-class of causal systems such that fixed-points can be completely characterised by the selection function $\texttt{Reach}_{\mathcal{M}}(w,\ E)$. To highlight relationships (originally presented in a previous work [45]) that exist among this sub-class, a sub-class of state elimination systems, and state transition systems based on $\texttt{Reach}_{\mathcal{M}}(w,\ E)$, we introduce the following definitions.

Definition 4.7.2 *($\mathcal{Q}$ Closed under disjunction)*
We shall say that a causal system $\mathcal{Q}$ is closed under disjunction *if and only if whenever $\varphi \Rightarrow \psi$ and $\chi \Rightarrow \zeta$ are in $\mathcal{Q}$ then $(\varphi \vee \chi) \Rightarrow (\psi \vee \zeta)$ is also in $\mathcal{Q}$.*

Definition 4.7.3 *($\mathcal{S}$ Closed under union)*
We shall say that a state elimination system $\mathcal{S}$ is closed under union *if and only if whenever $(X \rhd Y)$ and $(Q \rhd R)$ are in $\mathcal{S}$ then $(X \cup Q) \rhd (Y \cup R)$ is also in $\mathcal{S}$.*

The relationships are given then by the following results.

Theorem 4.7.4 *Every causal system closed under disjunction is selection-equivalent to a state elimination system closed under union. Conversely, every state elimination system closed under union is selection-equivalent to a causal system closed under disjunction.*

Theorem 4.7.5 *For every state elimination system $\mathcal{S}$ that is closed under union, there exists a selection-equivalent state transition system $\mathcal{M}$ based on* $\texttt{Reach}_{\mathcal{M}}(w,\ E)$. *Conversely, for every state transition system $\mathcal{M}$ based on* $\texttt{Reach}_{\mathcal{M}}(w,\ E)$ *there is a selection-equivalent state elimination system $\mathcal{S}$ that is closed under union.*

Note that disjunction in terms of sentences ($\phi \vee \psi$) corresponds to union in terms of states ($[\phi \vee \psi] = [\phi] \cup [\psi]$).

By combining theorems 4.7.4 and 4.7.5 the following corollary (analogous to Corollary 4.6.6) can be derived.

Corollary 4.7.6 *For every causal system $\mathcal{Q}$ closed under disjunction, there exists a selection-equivalent state transition system $\mathcal{M}$ based on* $\texttt{Reach}_{\mathcal{M}}(w,\ E)$. *Conversely, for every state transition system $\mathcal{M}$ based on* $\texttt{Reach}_{\mathcal{M}}(w,\ E)$ *there exists a selection-equivalent causal system $\mathcal{Q}$ which is closed under disjunction.*

Clearly, the causal system of the example considered in the previous section 4.7.1 is not closed under disjunction. That was the reason why the state transition system based on $\texttt{Reach}_{\mathcal{M}}(w,\ E)$ produced successor state(s) that were not causal fixed-points. It can be easily verified that if we close this causal system under disjunction, it will include the following causal rules in particular:

$$\neg c \ \Rightarrow\ \neg a \wedge \neg c,$$

$$\neg c \ \Rightarrow\ \neg b \wedge \neg c.$$

These two rules together with the original ones ensure that the state r, that is still selected by function $\texttt{Reach}_{\mathcal{M}}(w,\ \neg c)$, is now a causal fixed-point, $r \in Res_{\mathcal{Q}}(w,\ \neg c)$. It is worth pointing out that the state elimination system is now closed under union and includes the rule

$$\{p, q, r\} \rhd \{r\}.$$

In short, an augmented preferential semantics in terms of state transition systems based on the function $\texttt{Reach}_{\mathcal{M}}(w,\ E)$ characterises only those causal systems where the causal rules are closed under disjunction. Consequently, we needed a stronger selection

function $\texttt{Succ}_{\mathcal{M}}(w,\ E)$ to exactly capture all causal systems. To reiterate, the Hamiltonian path used in $\texttt{Succ}_{\mathcal{M}}(w,\ E)$, serves as a contextual mechanism allowing the domain causality to contribute only in certain situations.

4.7.3 Summary

It has been suggested that other preferential-style approaches to reasoning about action are capable of capturing McCain and Turner's causal theory of actions [20, 31, 63, 68]: an apparent contradiction of what is suggested by our impossibility theorems in Section 4.3. However, these and similar approaches[5] are able to do so only by augmenting the original language $\mathcal{B}$, with, for instance, predicates like *occludes* [20], *Caused* [33], or so-called inertial (frame) fluents [31]. In particular, the Logic of Universal Causation based on McCain and Turner action theories and introduced by Turner [68] has been shown to be remarkably similar to Lin's circumscriptive preferential action theories [33] that explicitly utilise the additional predicate *Caused*. In short, these approaches do not have a *pure preferential* semantics. At the outset we made clear that we did not wish to adopt the tactic of enhancing the underlying language. The ontological status of added predicates is not always clear and places a burden on the designer who must determine whether and when to occlude predicates. In the next chapter, however, we discuss a case when the original language $\mathcal{B}$ can be extended with extra elements that have a clear ontological status, while the semantics for selecting successor states remains transparent. In short, we hope to demonstrate that it may be possible to capture other causal approaches such as that of Thielscher [63] with our augmented preferential semantics. Another avenue we intend to explore in this work is a comparison between Sandewall's *causal propagation semantics* [56] and our semantics. This would link Sandewall's semantics with McCain and Turner's causal theory of actions and Thielscher's causal relationships approach, giving further insight into causal approaches to reasoning about action.

In summary, providing McCain and Turner's causal theory of action with an augmented preferential semantics allows comparison with other logics of action and makes a concrete step towards a unifying semantics.

[5] The action languages approach [31] was considered in the previous chapter, and the causal relationship approach [63] will be analysed in the next one.

Chapter 5

Causal Relationships Approach

In the previous chapter we presented a new semantics for McCain and Turner's approach. In this chapter we make a further step towards a unifying semantic framework for approaches to reasoning about action and change with an explicit causal component, and develop a semantics for an arguably more complex causal approach — that of Thielscher [63]. We begin by describing a semantics that, instead of concentrating on the standard state space, considers a larger set of possibilities—a *hyper-state space*— and traces the effects (both direct and indirect) with the states of the hyper-state space. In an intuitive sense, the states of the hyper-state space supply extra contextual information to track the presence of causality. We then present a further semantics which abstracts away certain important features of the hyper-state space approach. This *power-state space* semantics is a variant of the augmented preferential semantics, where power states stand for information states.

In the following section we outline the necessary technical preliminaries for an understanding of this chapter and describe Thielscher's causal theory of action. In Section 5.2 we describe the hyper-state space semantics that we shall use to characterise Thielscher's [63] approach. Then we introduce an abstraction of the hyper-state space semantics that we call the power-state space semantics. Section 5.4 will establish the necessary representation theorems. Section 5.5 discusses the importance of these results.

5.1 Technical Preliminaries and Background

In this section we review Thielscher's [63] causal relationships approach, and reproduce, for convenience, some of the technical preliminaries described in Chapter 2. We will adopt from Thielscher [63] the following notation. If $\epsilon \in L_{\mathcal{F}}$, then $|\epsilon|$ denotes its affirmative component, that is, $|f| = |\neg f| = f$, where $f \in \mathcal{F}$. This notation can be extended to sets of fluent literals as follows: $|S| = \{|f| : f \in S\}$. By the term *state* we intend a maximal consistent set of fluent literals. We will denote the set of all states as $\mathcal{W}$, and call the number m of fluent names in $\mathcal{F}$ the *dimension* of $\mathcal{W}$. By $[\phi]$ we denote all states consistent with the sentence $\phi \in \mathcal{B}$ (i.e., $[\phi] = \{w \in \mathcal{W} : w \vdash \phi\}$). Occasionally, we shall use $[E]$, where E is a set of literals, to denote all states consistent with a sentence $\wedge E$ (a conjunction of all literals in E). Domain constraints are sentences which have to be satisfied in all states.

We mentioned earlier that the idea of minimising change in order to deduce the set of possible next states (successor states) is used quite broadly in action theories. Sometimes the notion of minimal change is defined by set inclusion (eg., PMA) [70, 23, 31, 37], and often incorporates the frame concept or the policy of categorisation [23, 67, 31], assigning different degrees of inertia to language elements (fluents, literals, formulas, etc.) — as we have seen in action languages. Shortcomings of particular implementations of the Principle of Minimal Change and the policy of categorisation are well-known: imprecise or capricious definitions of minimality metrics (eg., PWA [17] vs PMA [70]), difficulties in properly categorising fluents as inertial and non-inertial, leading to increasingly complex selection mechanisms of action languages [31, 48, 63]), etc. These problems have generated attempts to use some notion of causality instead of, or in addition to, the Principle of Minimal Change. Action theories, discussed in previous chapters, try to embody background information in the form of domain "causal laws", pointing to the fact that, in general, propositions embracing causal dependencies are more expressive than traditional state constraints [33, 37, 67].

However, despite numerous attempts to combine a notion of causality with the Principle of Minimal Change and/or policy of categorisation, multiple counter-examples keep reappearing, highlighting the intractability of the Ramification problem. The frame-

work suggested by Thielscher [63] criticised the categorisation policy and the Principle of Minimal Change, arguing for the necessity of an approach based on causality. The suggested approach was intended to provide a method to avoid counter-intuitive indirect effects (ramifications), while accounting for causal relationships of a domain at hand. One of the perceived strengths of the Thielscher approach was an ability to capture not only all intuitively expected successor states with minimal distance to the initial state, but also non-minimal solutions — "perfectly acceptable provided all changes are reasonable from the standpoint of causality" [63]. In other words, the non-minimal solutions are those states which are reachable via causal propagation from an intermediate state. This intermediate state is determined as the nearest state to the initial state, where the direct effects of an action hold, while some domain constraints may be violated.

The assumption of minimal change, criticised by Thielscher [63], suggests to reject a successor state if it is obtained by changing the values of strictly more fluents than necessary (essentially, it is the PMA assumption). Arguably, this assumption is too restrictive and specific to warrant a complete withdrawal from the Principle of Minimal Change. Moreover, in this chapter we argue that it is possible to observe a minimal change component in Thielscher's approach as well. To demonstrate our claim we exhibit a semantics for Thielscher's causal theory of actions. This semantics is a variant of the augmented preferential semantics, and can be clearly seen to employ a component of minimal change coupled with causality.

5.1.1 Propagation with Causal Relationships

Thielscher's [63] causal theory of action consists of two main components: *action laws* which describe the direct effects of an action performed in a given state, and *causal relationships* which determine the indirect effects of an action.

Every action law consists of:

- a condition C, which is a set of fluent literals, all of which must be contained in an initial state where the action is intended to be applied;
- a (direct) effect E, which is also a set of fluent literals, all of which must hold in the resulting state after having applied the action.

Condition and effect are constructed from the same set of fluent names so that the state obtained from a direct effect is determined by removing C from the initial state and adding E to the result. An action may result in a number of state transitions.

Definition 5.1.1 *(Action)*
Let $\mathcal{F}$ be the set of fluent names and let $\mathcal{A}$ be a finite set of symbols called action names, such that $\mathcal{F} \cap \mathcal{A} = \emptyset$. An action law is a triple $\langle C, a, E \rangle$ where C, called condition*, and E, called* effect*, are individually consistent sets of fluent literals, composed of the very same set of fluent names (i.e., $|C| = |E|$) and $a \in \mathcal{A}$. If w is a state then an action law $\alpha = \langle C, a, E \rangle$ is applicable in w if and only if $C \subseteq w$. The application of α to w yields the state $(w \setminus C) \cup E$ (where $\setminus$ denotes set subtraction).*

Causal relationships are specified as

$$\epsilon \ \textit{causes} \ \rho \ \textit{if} \ \Phi$$

where ϵ and ρ are fluent literals and Φ is a fluent formula based on the set of fluent names $\mathcal{F}$. Thielscher also proposed a mechanism to extract causal relationships from domain constraints, given a binary relation $\mathcal{I}$ indicating potential *influence* dependencies between certain fluents. This section would be described in more detail in [63].

Definition 5.1.2 *(Causal relationship applicability)*
Let (s, E) be a pair consisting of a state s and a set of fluent literals E. Then a causal relationship ϵ causes ρ if Φ is applicable to (s, E) if and only if $\Phi \wedge \neg\rho$ is true in s, and $\epsilon \in E$. Its application yields the pair (s', E'), denoted as $(s, E) \rightsquigarrow (s', E')$, where $s' = (s \setminus \{\neg\rho\}) \cup \{\rho\}$ and $E' = (E \setminus \{\neg\rho\}) \cup \{\rho\}$.

In other words, a causal relationship is applicable if Φ holds, the indirect effect ρ is false and the cause ϵ is among the current effects. Note that ϵ must be among the current effects; being true at the current state is not sufficient.

A possible *successor state* is determined through the repeated application of causal relationships. In so doing we may temporarily end up in states violating domain constraints. This is permissible provided subsequent applications of causal laws result in legal states. Specifically, given an initial state w and action a, the set of possible successor states $Res_{RD\mathcal{L}}(w, a)$ is determined as follows.

Definition 5.1.3 **($Res_{RD\mathcal{L}}(w, a)$)**
Let $\mathcal{F}$ be the set of fluent names, A a set of action names, $\mathcal{L}$ a set of action laws, D a set of domain constraints, and R a set of causal relationships. Furthermore, let w be a state satisfying D and let $a \in A$ be an action name. A state r is a successor state *of w and a, denoted $r \in Res_{RD\mathcal{L}}(w, a)$, if and only if there exists an applicable (with respect to w) action law $\alpha = \langle C, a, E\rangle \in \mathcal{L}$ such that*

1. $((w \setminus C) \cup E, E) \overset{*}{\rightsquigarrow} (r, E')$ *for some E', and*
2. *r satisfies D,*

where $\overset{*}{\rightsquigarrow}$ denotes the transitive closure of $\rightsquigarrow$.

As mentioned before, an occurrence of a literal ϵ in a state s does not guarantee that a causal relationship ϵ *causes* ρ *if* Φ is applicable to a pair (s, E) — to ensure applicability, the literal ϵ has to belong to the current effects E. That is why, in order to trace causal propagation with causal relationships, one needs to keep an explicit (and changing) account of context-dependent effects of actions.

5.1.2 The Light Detector Example

At this stage let us consider the Light Detector example, mentioned earlier, and illustrate the way of producing successor states according to the causal relationship approach. The example can be characterised as follows:

$$\mathcal{F} = \{sw_1, sw_2, sw_3, relay, light, detect\}$$

$$D = \{sw_1 \wedge sw_2 \leftrightarrow light,\ \ sw_1 \wedge sw_2 \leftrightarrow relay,$$

$$relay \supset \neg sw_2,\ \ light \supset detect\}$$

$$R = \{sw_1 \textit{ causes } light \textit{ if } sw_2,\ \ sw_2 \textit{ causes } light \textit{ if } sw_1,$$

$$\neg sw_1 \textit{ causes } \neg light \textit{ if } \top,\ \ \neg sw_2 \textit{ causes } \neg light \textit{ if } \top,$$

$$sw_1 \textit{ causes } relay \textit{ if } sw_3,\ \ sw_3 \textit{ causes } relay \textit{ if } sw_1,$$

$$\neg sw_1 \textit{ causes } \neg relay \textit{ if } \top,\ \ \neg sw_3 \textit{ causes } \neg relay \textit{ if } \top,$$

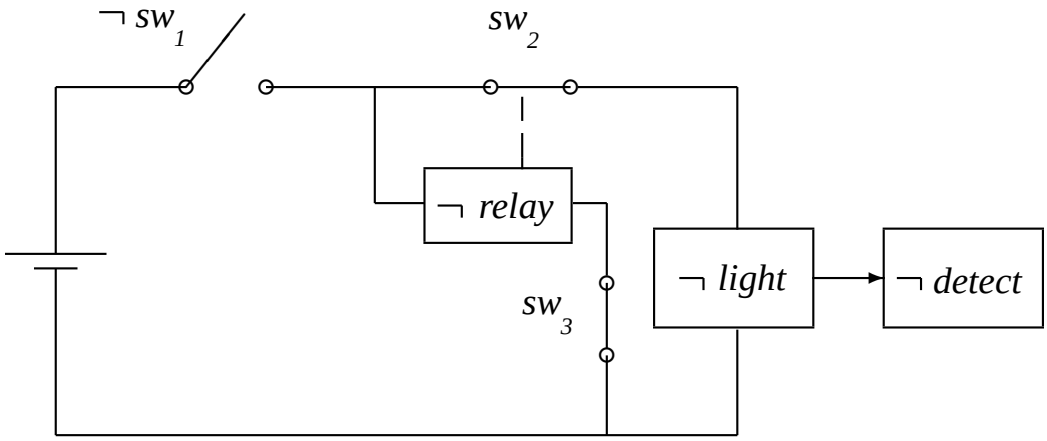

Figure 5.1: The electric circuit with Light Detector.

$$relay \textit{ causes } \neg sw_2 \textit{ if } \top,\ light \textit{ causes } detect \textit{ if } \top\}$$

Let the initial state be

$$w = \{\neg sw_1, sw_2, sw_3, \neg relay, \neg light, \neg detect\},$$

depicted in Figure 5.1.2. Consider now the action $\langle\{\neg sw_1\}, toggle_1, \{sw_1\}\rangle$. According to Thielscher [63, p. 341],

> Obviously, the relay gets activated and, then, attracts the second switch, sw_2. Hence, the latter is open in the finally resulting state. Notice, however, that as soon as the first switch is closed, the sub-circuit involving the light bulb gets closed, too. This may activate the light bulb for an instant, that is, before the second switch jumps its position as a result of activating the relay. If this is indeed the case, then this short-time activation might be registered by the photo device, *detect*. Hence, while it is clear that the light bulb is off in the resulting state, it may or may not be the case that *detect* becomes true.

Accordingly, two successor states are suggested as intuitive depending on current flow in the circuit:

$$s_1 = \{sw_1, \neg sw_2, sw_3, relay, \neg light, \neg detect\},$$

$$s_2 = \{sw_1, \neg sw_2, sw_3, relay, \neg light, detect\}.$$

Importantly, the second state s_2 differs from the initial state w in strictly more fluents than state s_1 — in other words, $s_1 \prec_w s_2$ in terms of the PMA ordering.

The set R of causal relationships supports this line of reasoning completely. The execution of $\langle\{\neg sw_1\}, toggle_1, \{sw_1\}\rangle$ in the state w produces the state-effect pair

$$(\{sw_1, sw_2, sw_3, \neg relay, \neg light, \neg detect\}, \{sw_1\}).$$

Then, by applying various causal relationships, one may reach the following state-effect pairs:

$$(s_1, \{sw_1, relay, \neg sw_2\}),$$

$$(s_1, \{sw_1, relay, \neg sw_2, \neg light\}),$$

$$(s_2, \{sw_1, detect, relay, \neg sw_2, \neg light\}).$$

The last pair includes the alternative outcome s_2 because during the propagation it is possible to employ the rules

$$sw_1 \ \textit{causes} \ light \ \ if \ \ sw_2$$

$$light \ \textit{causes} \ detect \ \textit{if} \ \top$$

before the rules

$$sw_1 \ \textit{causes} \ relay \ \textit{if} \ sw_3$$

$$relay \ \textit{causes} \ \neg sw_2 \ \textit{if} \ \top$$

$$\neg sw_2 \ \textit{causes} \ \neg light \ \textit{if} \ \top.$$

In other words, the literal $light$ appears in the effects (history) component of some state-effect pair, and brings about the literal $detect$, before being removed itself by other rules, while the ramification $detect$ stays.

Thus, both successor states s_1 and s_2 are obtained by the causal relationships approach in the Light Detector example — at the expense of keeping an explicit account of all intermediate contextual changes (that are not necessarily retained in successor states).

5.1.3 Preliminary Comments

Interestingly, given a *transition* state-effect pair (s, E), if the literal ϵ is part of the current effects E, then it must be an element of the current state s. This observation can be formalised as follows.

Lemma 5.1.4 *If* $(s', E') \overset{*}{\rightsquigarrow} (s'', E'')$, *then* $E'' \subseteq s''$.

Notice that the set E' in Definition 5.1.3 contains the most recent consistent effects that have been manifested during the causal propagation $((w \setminus C) \cup E, E) \overset{*}{\rightsquigarrow} (r, E')$. In other words, although some of the effects may have been retracted from the effects set during propagation, their negations should have taken their respective places. The effects set accounts for both direct and indirect effects but they are not necessarily preserved during causal propagation.

For example, suppose we have a simple action system with $\mathcal{F} = \{p, q\}$, $D = \{\neg q \supset \neg p\}$, $R = \{\neg q \ \textit{causes} \ \neg p \ \textit{if} \ \top\}$, and $\mathcal{L} = \{\langle\{p, q\}, a, \{p, \neg q\}\rangle\}$. The action a performed at the initial state $\{p, q\}$, results in a state $\{p, \neg q\}$. Note that this resultant state does not satisfy the domain constraint. The causal relationship is then applied, whereby $(\{p, \neg q\}, \{p, \neg q\}) \rightsquigarrow (\{\neg p, \neg q\}, \{\neg p, \neg q\})$ and produces $Res_{RD\mathcal{L}}(\{p, q\}, a) = \{\neg p, \neg q\}$, where the successor state satisfies D, while leaving one of the direct effects (p) out.

We can strengthen the concept of successor states to *conservative* successor states. A stronger definition describing *conservative successor states* (denoted $Res^*_{RD\mathcal{L}}(w, a)$) can be given as follows.

Definition 5.1.5 ($Res^*_{RD\mathcal{L}}(w, a)$)
Let $\mathcal{F}$, A, $\mathcal{L}$, D, R, w, $\alpha = \langle C, a, E\rangle$ *be the same as in Definition 5.1.3. A state* r *is a* conservative successor state *of* w *and* a, $r \in Res^*_{RD\mathcal{L}}(w, a)$, *if and only if*

1. $r \in Res_{RD\mathcal{L}}(w, a)$, *and*
2. $E \subseteq r$.

Using this definition, causal propagation can "travel" outside E-states, however it must end in a state consistent with the direct effects E.

Our primary intention is to characterise a successor state r in terms of initial state w and action e, without referring explicitly to the details of the intermediate states. In other words, it is desirable to define a selection function $Res(w, a)$ that does not trace a detailed history of intermediate effects. In doing so, we shall follow the spirit of the augmented preferential semantics described in Chapter 2.

5.2 Hyper-space Semantics

In Chapter 2 we discussed a semantics that augmented a preferential structure with a binary relation on states, and argued that minimal change and causality — the former captured by preferential semantics and the latter by a binary relation — together are essential in furnishing a concise solution to the frame problem. Our approach here is intended to illustrate this idea once more, now with respect to Thielscher's causal theory of action.

This will serve as another step towards a uniform and general augmented preferential semantics for causal action systems — complementing the Principle of Minimal Change[1] with causal propagation driven solely by a binary relation on states.

Our intention at this stage is to consider a formalisation of action systems which faithfully captures successor states (not only conservative), as defined by $Res_{RD\mathcal{L}}(w, a)$ or $Res^*_{RD\mathcal{L}}(w, a)$, using a selection mechanism of the augmented preferential semantics (or its variant). More precisely, we would like to use a binary (causal) relation on states, while propagating within a set of possible states, instead of keeping an explicit (and changing) account of context-dependent effects of actions. The advantage of this proposal is that a causal relation would be action-independent, unlike the history of effects. Obviously, this objective is hardly achievable without extending the action systems components in some way — in particular, we cannot use the approximation $\mathcal{W} = \Gamma$.

In general, as we discussed earlier, a pure preferential semantics, in the spirit of [60], cannot be obtained for causal action systems without extending the system description. One important feature of such an extension is the introduction of the information state-space Γ with a (typically) higher dimension than the standard state-space $\mathcal{W}$.

[1] Some notion of minimality was used, for instance, in obtaining the state $(w \setminus C) \cup E$.

Let us begin by informally describing the semantics we develop for the causal relationship approach, before proceeding to establish the formal results. In the remainder of this section we give a formal description of this semantics (a variant of the augmented preferential semantics). This variant makes use of the information state-space Γ explicitly, by constructing it in two steps — first, via the "hyper-state space" (following our original techniques [50]), and then via the "power-state space" (introduced by us earlier in [51]).

Any state in the standard state-space $\mathcal{W}$ can be associated with a number of hyper-states, creating a hyper-neighbourhood. For instance, an intermediate state $(w \setminus C) \cup E$ (defined, for a given action and an initial state, according to Thielscher's approach) can be represented by a set of hyper-states in the expanded space. This hyper-neighbourhood will be a starting point of a propagation (analogously to the gradient area in the augmented preferential semantics[2]). An appropriately constructed binary relation on hyper-states would allow us to propagate in the hyper-space in a very simple way — without the necessity to track the causal history, and resulting in a clearly defined "final" set of hyper-states. A projection from the resulting hyper-neighbourhood back to the normal state-space would pin-point the desired successor state of the action at hand — see Figure 5.2. Intuitively, the purpose of the hyper-states is to serve as possible causal extensions of normal states, providing necessary context to the process of causal propagation.

5.2.1 Constructing Hyper-states

The main idea behind the hyper-state space semantics is to construct a prototype of the information state-space and investigate the applicable forces of minimality and causality. Technically, we do this by augmenting the underlying language and adding to the set of fluent names $\mathcal{F}$. This allows us to maintain contextual information that will become important during causal propagation.

We begin by considering the set $\overset{\circ}{\mathcal{F}}$ which has the same cardinality as $\mathcal{F}$ and such that $L_{\mathcal{F}} \cap \overset{\circ}{\mathcal{F}} = \emptyset$. Then we define a function $j : \mathcal{F} \to \overset{\circ}{\mathcal{F}}$ which intuitively provides an added space-dimension corresponding to each fluent $f \in \mathcal{F}$. Now consider the set

[2]This analogy is limited, however: the hyper-state space is only an initial prototype of the information space.

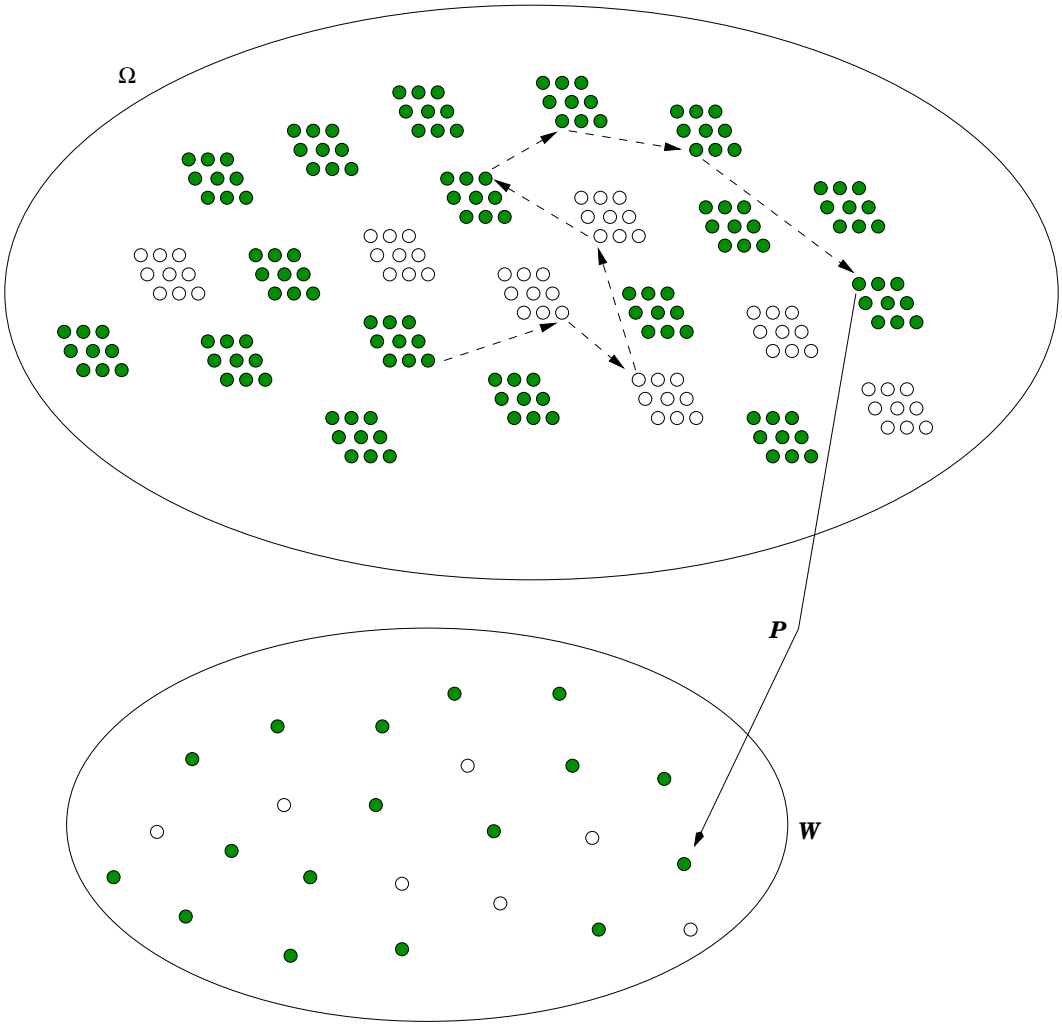

Figure 5.2: Propagation in the hyper-state space (dashed arrows), followed by projection onto the normal state-space (solid arrow).

$\overset{\circ}{L_{\mathcal{F}}} = \overset{\circ}{\mathcal{F}} \cup \{\neg q : q \in \overset{\circ}{\mathcal{F}}\}$. Clearly, $\overset{\circ}{L_{\mathcal{F}}}$ and $L_{\mathcal{F}}$ are sets of the same cardinality and, moreover, $L_{\mathcal{F}} \cap \overset{\circ}{L_{\mathcal{F}}} = \emptyset$. Put simply, we just double the number of fluents (and literals), by adding their "copies".

We require a further function to map from $L_{\mathcal{F}}$ to $\overset{\circ}{L_{\mathcal{F}}}$. We introduce the "cloning" function $l : L_{\mathcal{F}} \rightarrow \overset{\circ}{L_{\mathcal{F}}}$ for this purpose such that

$$l(f) = j(f) \text{ if } f \in \mathcal{F}; \ f \text{ is a positive literal — a fluent name, and}$$

$$l(f) = \neg j(|f|) \text{ if } f \in L_{\mathcal{F}} \setminus \mathcal{F}; \ f \text{ is a negative literal.}$$

In other words, the function $l(f)$ is a simple bijection mapping literals from $L_{\mathcal{F}}$ to $\overset{\circ}{L_{\mathcal{F}}}$: positive to positive, and negative to negative.

The following property follows from these definitions.

Lemma 5.2.1 *If $f \in \mathcal{F}$, then $l(\neg f) = \neg l(f)$.*

The function $l(f)$ is intended to produce extra literals, corresponding to fluent literals in $L_{\mathcal{F}}$. We will call a literal $l(f)$ a justifier literal, and will use the abbreviation $\overset{\circ}{f}$ instead of $l(f)$ for simplicity. In addition, the set $\overset{\circ}{\mathcal{F}}$ will be referred to as the set of *justifier fluents*, and $\overset{\circ}{L_{\mathcal{F}}}$ will be referred to as the set of *justifier literals*.

Definition 5.2.2 *(Justifier set)*
A justifier set $\overset{\circ}{J}$*, for a set of fluent literals J, is $\overset{\circ}{J} = \cup_{f \in J}\{l(f)\} = \cup_{f \in J}\{\overset{\circ}{f}\}$.*

We are now in a position to state more precisely what we mean by a hyper-state.

Definition 5.2.3 *(Hyper-state)*
Given a set of fluents $\mathcal{F}$, a hyper-state *is a maximal consistent set of literals from $L_{\mathcal{F}} \cup \overset{\circ}{L_{\mathcal{F}}}$.*

That is, we produce "clones" or copies of all the fluent names in our language and use this expanded language in forming (hyper-)states. We will denote the set of all hyper-states as Ω, where the dimension of Ω is $2m$, m being the dimension of $\mathcal{W}$. The following two functions map hyper-state space Ω to normal space $\mathcal{W}$ and vice versa.

Definition 5.2.4 *(Projection from hyper-state space)*
A projection from Ω to $\mathcal{W}$, $p : \Omega \to \mathcal{W}$, is the function mapping a hyper-state $s = \{f_1, ..., f_m, \overset{\circ}{f}_1, ..., \overset{\circ}{f}_m\} \in \Omega$ to a state $r = \{f_1, ..., f_m\} \in \mathcal{W}$.

We denote the hyper-part of a hyper-state $s \in \Omega$ as $h(s) = s \setminus p(s)$. Clearly, for any $s \in \Omega$, $h(s) \cap \mathcal{F} = \emptyset$.

The following definition will be useful in providing a semantics for the process of causal propagation.

Definition 5.2.5 *(Hyper-neighbourhood)*
A hyper-neighbourhood *of a state $r \in \mathcal{W}$ is the function $N : \mathcal{W} \to 2^{\Omega}$, mapping a state r to a set of hyper-states: $N(r) = \{s \in \Omega : r = p(s)\}$.*

That is, the hyper-neighbourhood $N(r)$ is the set of all hyper-states extending r (consistent with r).

5.2.2 Justifying Causal Context

Clearly, there are 2^m states in any hyper-neighbourhood, as there are m justifier fluent names in any hyper-state allowed to vary across the neighbourhood. Intuitively, justifier literals represent explicit causes for a state $r \in \mathcal{W}$, and the set $N(r)$ consists of states where all possible causes (i.e., justifier literals) vary, while the (proper) literals defined on $\mathcal{F}$ are fixed.

For instance, given state $r = \{a, b\}$ in normal space $\mathcal{W}$, its hyper-neighbourhood $N(r)$ consists of hyper-states

$$\{a, b, \overset{\circ}{a}, \overset{\circ}{b}\},$$
$$\{a, b, \overset{\circ}{a}, \neg\overset{\circ}{b}\},$$
$$\{a, b, \neg\overset{\circ}{a}, \overset{\circ}{b}\},$$
$$\{a, b, \neg\overset{\circ}{a}, \neg\overset{\circ}{b}\},$$

where the justifier fluents $\overset{\circ}{a}$ and $\overset{\circ}{b}$ vary. As such, any subset of $N(r)$ may represent a particular causal context. For example, the set $\{\{a, b, \overset{\circ}{a}, \overset{\circ}{b}\}, \{a, b, \overset{\circ}{a}, \neg\overset{\circ}{b}\}\}$ corresponds to a partial state $\{a, b, \overset{\circ}{a}\}$, justifying the literal $a \in r$, and leaving the literal $b \in r$

unsupported (more precisely, any *change* in truth value of a literal will be expected to have a justification).

Note that the history component E of any causally propagated pair (s, E) cannot have more than m elements due to the consistency of the update defined in Definition 5.1.2, as shown by Lemma 5.1.4. In the case where the history component E in a pair (s, E) has exactly m elements (or, in other words, $E = s$, by the Lemma 5.1.4) the pair can be represented by a single hyper-state $s \cup \overset{\circ}{s}$. For example, the hyper-state $\{a, b, \overset{\circ}{a}, \overset{\circ}{b}\}$ can account for a causal transition pair $(\{a, b\}, \{a, b\})$. When the component E has strictly fewer elements, $E \subset s$, the incompleteness may be represented by a partial hyper-state. A union of complete hyper-states, $\{\{a, b, \overset{\circ}{a}, \overset{\circ}{b}\}, \{a, b, \overset{\circ}{a}, \neg\overset{\circ}{b}\}\}$ can represent the pair $(\{a, b\}, \{a\})$ in a causal propagation chain where the second component carries the history of change $\{a\}$.

It is this combinatorial variability of possible causes in a hyper-neighbourhood that allows us to account for different action-dependent histories in a causally propagated chain, leading to a successor state in $Res_{RD\mathcal{L}}(w, a)$. Before we formally introduce the required notion of a binary causal relation on hyper-states, let us illustrate its purpose.

Suppose we have an action system with $\mathcal{F} = \{a, b, c\}$, $D = \{\neg b \supset \neg a\}$, $R = \{\neg b$ *causes* $\neg a$ *if* $\top\}$, and $\mathcal{L} = \{\langle\{b\}, x, \{\neg b\}\rangle\}$. Let us consider action x executed in the initial state $w = \{a, b, c\}$. The action's direct effect is $\{\neg b\}$, yielding the intermediate state $\{a, \neg b, c\} = (w \setminus \{b\}) \cup \{\neg b\}$. This state contradicts the given domain constraint. However, the system's sole causal law applies: $(\{a, \neg b, c\}, \{\neg b\}) \rightsquigarrow (\{\neg a, \neg b, c\}, \{\neg a, \neg b\})$. The state component of the resultant pair obeys the domain constraint (and satisfies the direct effect, in addition). Therefore, it is an element of $Res_{RD\mathcal{L}}(w, x)$. It can be verified that $Res_{RD\mathcal{L}}(w, x)$ is a singleton.

We now indicate how this propagation can be traced in the hyper-state space. The hyper-neighbourhood $N(q)$ of the intermediate state $q = \{a, \neg b, c\}$ contains eight hyper-states (see Figure 5.3). Some of these represent the initial history component $\{\neg b\}$ — these hyper-states are exactly those in $N(q) \cap [\neg\overset{\circ}{b}]$, where $[f]$ represents, as usual, the set of states consistent with literal f. The hyper-neighbourhood of the successor state $q' = \{\neg a, \neg b, c\}$ contains some hyper-states accountable for the final history component $\{\neg a, \neg b\}$. These states are exactly those in $N(q') \cap [\neg\overset{\circ}{a} \wedge \neg\overset{\circ}{b}]$ or $N(q') \cap [\neg\overset{\circ}{a}] \cap [\neg\overset{\circ}{b}]$. The

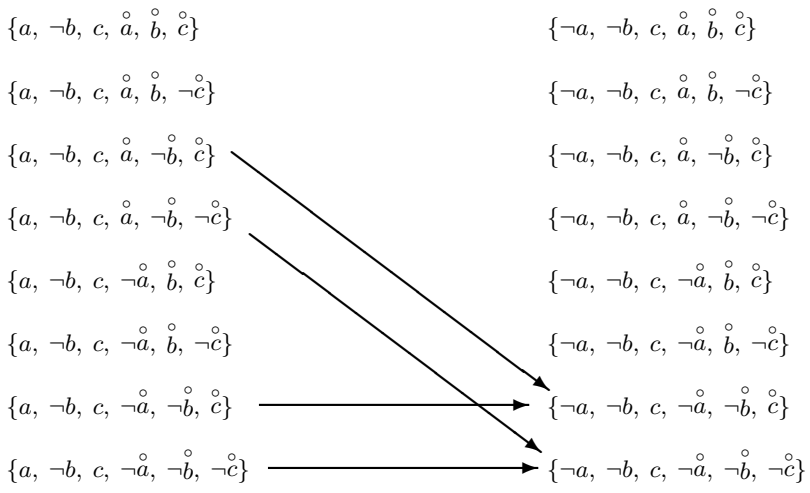

Figure 5.3: The $\mathcal{C}$-links between hyper-neighbourhoods of the states $\{a, \neg b, c\}$ and $\{\neg a, \neg b, c\}$, generated by a causal relationship $\neg b$ *causes* $\neg a$ *if* $\top$.

idea, then, is to construct just such a binary relation on hyper-states for an action system so that transitions in hyper-state space correspond to causal propagation.

We now can formally define a binary relation on hyper-states in Ω.

Definition 5.2.6 *(Binary relation $\mathcal{C}$)*
A binary relation $\mathcal{C}$ is defined on $\Omega \times \Omega$. We say that $\mathcal{C}(s, s')$ if and only if there exists a causal relationship ϵ causes ρ if Φ such that

1. $p(s) \vdash \epsilon \wedge \Phi \wedge \neg\rho$
2. $h(s) \vdash \mathring{\epsilon}$
3. $p(s') = (p(s) \setminus \{\neg\rho\}) \cup \{\rho\}$
4. $h(s') = (h(s) \setminus \{\neg\mathring{\rho}\}) \cup \{\mathring{\rho}\}$

That is,

1. the causal relationship is applicable at the world state $p(s)$ and ρ will change value
2. the antecedent ϵ of the causal relationship is among current effects — and therefore, the justifier $\mathring{\epsilon}$ is an element of the hyper-state s

3. the state $p(s')$ is like $p(s)$ but ρ has changed value

4. the effect ρ is added to current effects — and therefore, the justifier $\overset{\circ}{\rho}$ is an element of the hyper-state s'.

Figure 5.3 illustrates $\mathcal{C}(s, s')$-links between hyper-states from distinct hyper- neighbourhoods. The fact that all the states in $N(r) \cap [\neg\overset{\circ}{b}]$ have $\mathcal{C}$-links to the states in $N(r') \cap [\neg\overset{\circ}{a} \wedge \neg\overset{\circ}{b}]$ is not a coincidence, and will be formally captured later.

5.3 Power-space Semantics

While the hyper-state space is a powerful concept that allows us to completely characterise Thielscher's approach, we shall now introduce another concept that abstracts away some of the important elements of the hyper-state space semantics. This notion of a so-called *power-state space* allows us to give a semantics that concentrates more on the actual causal propagation occurring between one (hyper-)neighbourhood and another. Moreover, this semantics is a simple variant of the augmented preferential semantics. First, however, the following definition will prove convenient.

Definition 5.3.1 *(Partial state)*
Given a state $q \in \mathcal{W}$ and a set $z \subseteq N(q)$, a partial state $\gamma_q(z)$ is defined as $\cap_{s \in z} s$.

Intuitively, a partial state $\gamma_q(z)$ contains all literals common to the "part" z of the neighbourhood $N(q)$. For example, the state $q = \{a, \neg b, c\}$ and the set $z = N(q) \cap [\neg\overset{\circ}{b}]$ yield a partial state $\gamma_q(z) = \{a, \neg b, c, \neg\overset{\circ}{b}\}$.

Consider a set Γ of cardinality equal to that of 2^Ω. This set Γ will be referred to as the *power-state space*, being isomorphic to the power set of the hyper-state space Ω.

We define a mapping $\gamma : 2^\Omega \rightarrow \Gamma$, such that $\gamma(z) = \gamma_q(z)$ if $z \subseteq N(q)$ for some $q \in \mathcal{W}$, and $\gamma(z) = \emptyset$ otherwise. Basically, if z is a part of some hyper-neighbourhood $N(q)$, the function $\gamma(z)$ is nothing but $\gamma_q(z)$. Otherwise, if z contains hyper-states from different hyper-neighbourhoods, $\gamma(z)$ is defined as the empty set.

Having defined the function $\gamma(z)$ for every subset of the hyper-state space Ω, we construct a binary relation on elements of Γ.

Definition 5.3.2 *(Binary relation $\rightharpoonup$)*
A binary relation $\rightharpoonup$ is defined on $\Gamma \times \Gamma$. Given two elements $x_1, x_2 \in \Gamma$ such that $x_1 \neq \emptyset$ and $x_2 \neq \emptyset$, we say that $x_1 \rightharpoonup x_2$ if and only if $x_1 = \gamma(z_1)$ and $x_2 = \gamma(z_2)$ for some $z_1, z_2 \in 2^{\Omega}$ such that

1. $\forall\, s \in z_1, \exists s' \in z_2$, *such that* $\mathcal{C}(s, s')$,

2. $\forall\, s' \in z_2, \exists s \in z_1$, *such that* $\mathcal{C}(s, s')$

We will abbreviate $x_1 \rightharpoonup x_2$, where $x_1 = \gamma(z_1)$ and $x_2 = \gamma(z_2)$, as $\gamma(z_1) \rightharpoonup \gamma(z_2)$. Intuitively, $\gamma(z_1) \rightharpoonup \gamma(z_2)$ means that there are no hyper-states in z_1 without an outgoing $\mathcal{C}$-link to some hyper-state in z_2, and there are no hyper-states in z_2 without an incoming $\mathcal{C}$-link from some hyper-state in z_1.

It is precisely this binary relation that captures causal propagation in Thielscher's system. To reiterate, the power-state space is at a much better level of abstraction than the hyper-state space, while we find that the latter is more convenient in terms of establishing the initial motivation and in proving the necessary results. Figure 5.4 shows how the hyper-state and power-state spaces are related, and Figure 5.5 exemplifies that $\gamma(N(q) \cap [\neg \overset{\circ}{b}]) \rightharpoonup \gamma(N(q') \cap [\neg \overset{\circ}{a} \wedge \neg \overset{\circ}{b}])$.

By $\overset{*}{\rightharpoonup}$ we denote the transitive closure of the binary relation $\rightharpoonup$.

5.4 Representation Theorems

We intend to demonstrate that it is possible to correctly capture successor states obtained by the causal relationship approach by causal propagation in the power-state space Γ driven by the binary relation $\rightharpoonup$. Such a process starts in a power-state that is minimal (in a certain way) among power-states consistent (in a certain way) with an action's direct effects, and ends in a final power-state corresponding to a successor state. In other words, this process propagates "minimal change" within a set of possible states of higher dimensions, instead of keeping an explicit (and changing) account of context-dependent action effects. Quite obviously, the power-space semantics is a variant of the augmented preferential semantics.

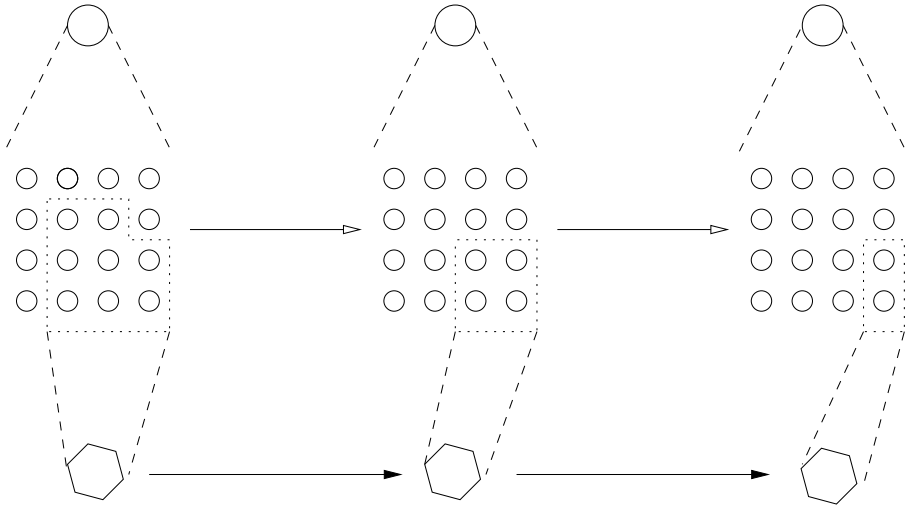

Figure 5.4: The top level: states in normal space $\mathcal{W}$. The medium level: states in hyper-state space Ω grouped in hyper-neighbourhoods corresponding to normal states. A subset of a hyper-neighbourhood (a partial hyper-state) is shown bounded by a dotted line. The bottom level: states in power-state space Γ corresponding to partial hyper-states.

5.4.1 Relating Justifier Literals and Causes

The aim of this section is to establish some representation conditions for the causal links $\mathcal{C}(s,\ s')$ in terms of the underlying causal relationships. This would allow us to imitate a chain of causal relationships with a propagation along the binary relation $\mathcal{C}$ and the binary relation $\rightharpoonup$.

We begin by introducing a few notions that will be useful in analysing causal links $\mathcal{C}(s,\ s')$ and $x \rightharpoonup x'$. Let $\mathcal{C}^*$ denote the transitive closure of $\mathcal{C}$.

Definition 5.4.1 *(Connection set)*
Given two states $x \in \mathcal{W}$ and $y \in \mathcal{W}$, the set $L(x,y) = \{s \in N(x) : \mathcal{C}(s,s')$, for some $s' \in N(y)\}$ is the connection set *for the states x and y.*

Intuitively, the connection set $L(x,y)$ contains those states from the hyper-neighbourhood $N(x)$ that have out-going $\mathcal{C}(s,s')$-links to some states in the hyper-neighbourhood $N(y)$. Note that, in general, $L(x,y) \neq L(y,x)$.

Definition 5.4.2 *(Traced set)*
Given two states $x \in \mathcal{W}$ and $y \in \mathcal{W}$, the set $T(x,y) = \{s' \in N(y) : \mathcal{C}(s,s')$, for some

$s \in N(x)\}$ *is the* traced set *for the states* x *and* y*.*

The traced set $T(x, y)$ contains those states from the hyper-neighbourhood $N(y)$ that have incoming $\mathcal{C}(s, s')$-links from some states in the hyper-neighbourhood $N(x)$.

Definition 5.4.3 *(Transitively traced set)*
Given two states $x \in \mathcal{W}$ *and* $y \in \mathcal{W}$*, the set* $T^*(x, y) = \{s' \in N(y) : \mathcal{C}^*(s, s')$*, for some* $s \in N(x)\}$ *is the* transitively traced set *for the states* x *and* y*.*

The transitively traced set $T^*(x, y)$ contains those states from the hyper-neighbourhood $N(y)$ that have are transitively reachable via incoming $\mathcal{C}^*(s, s')$-links from some states in the hyper-neighbourhood $N(x)$.

The relation $\mathcal{C}$ has a notable characteristic that there are at least 2^{m-1} links generated by one causal relationship ϵ *causes* ρ *if* Φ; among different hyper-states and neighbourhoods. The minimum 2^{m-1} is attained when the causal relationship is qualified by a complete state $r = \{f_1, \ldots, f_m\}$, where $\Phi \leftrightarrow \wedge_{k=1}^{k=m} f_k$ — in this case, all the generated links originate in one hyper-neighbourhood $N(r)$. This leads us to the following result.

Lemma 5.4.4 *For any two states* $x \in \mathcal{W}$ *and* $y \in \mathcal{W}$*, if the connection set* $L(x, y) \neq \emptyset$ *then there exists a justifier literal* $\overset{\circ}{f}$ *such that* $[\overset{\circ}{f}] \cap N(x) \subseteq L(x, y)$*.*

Essentially, this result tells us that, if there is at least one $\mathcal{C}$-link between two hyper-states, then there are at least 2^{m-1} $\mathcal{C}$-links in total between hyper-states in the respective neighbourhoods, and all these links are generated by the same causal relationship. As an example, Figure 5.5 illustrates the existence of a justifier literal $\neg\overset{\circ}{b}$ such that $[\neg\overset{\circ}{b}] \cap N(\{a, \neg b, c\}) \subseteq L(\{a, \neg b, c\}, \{\neg a, \neg b, c\})$.

A qualified reverse inclusion holds also.

Lemma 5.4.5 *For any two states* $x \in \mathcal{W}$ *and* $y \in \mathcal{W}$*, if there exists a justifier literal* $\overset{\circ}{f}$ *such that* $[\overset{\circ}{f}] \cap N(x) \subseteq L(x, y)$*, then there exists a causal relationship* f causes ρ if Φ*, for some* Φ *true in* x*, where* $\{\rho\} = y \setminus x$*.*

The proof of this lemma progressively eliminates all literals except f, which might have been alternative causes. It exploits the fact that varying $m - 1$ justifier literals

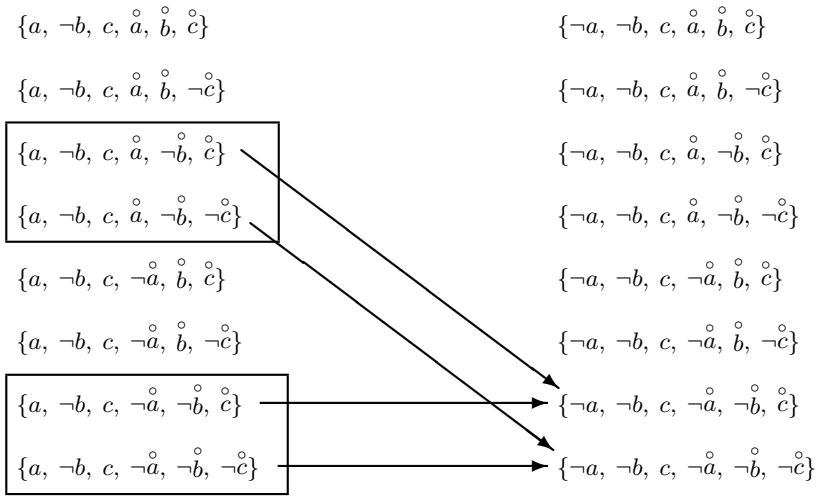

Figure 5.5: All $[\neg\overset{\circ}{b}]$-states belong to the connection set $L(\{a, \neg b, c\}, \{\neg a, \neg b, c\})$.

(having fixed $\overset{\circ}{f}$) accounts for at most $2^{m-1} - 1$ states in a hyper-neighbourhood, while there are 2^{m-1} states in the set $[\overset{\circ}{f}] \cap N(x)$.

Combining the previous two results shows that the presence of a causal relationship underlying a $\mathcal{C}$-link corresponds to the existence of a justifier literal $\overset{\circ}{f}$ such that $[\overset{\circ}{f}] \cap N(x) \subseteq L(x, y)$. More precisely,

Corollary 5.4.6 *For any two states $x \in \mathcal{W}$ and $y \in \mathcal{W}$, there exists a justifier literal $\overset{\circ}{f}$ such that $[\overset{\circ}{f}] \cap N(x) \subseteq L(x, y)$, if and only if there exists a causal relationship f* causes *ρ* if *Φ, for some Φ true in x, where $\{\rho\} = y \setminus x$.*

This representation result illustrates the mechanics behind the causal links $\mathcal{C}$. In particular, in order to imitate a chain of causal relationships with the binary relation $\mathcal{C}$, one would need to "group" those hyper-states that share a particular justifier literal and propagate to another "group" (in another hyper-neighbourhood) that shares both the original justifier literal and, in addition, the justifier of its causal effect.

Not surprisingly, any connection set may not contain all $[\overset{\circ}{\epsilon}]$-states and all $[\neg\overset{\circ}{\epsilon}]$-states in a hyper-neighbourhood. The set of causal relationships R may have relationships like *ϵ causes ρ if Φ* and *$\neg\epsilon$ causes ρ if Φ*. However, such "conflicting" relationships would generate $\mathcal{C}$-links originating from different hyper-neighbourhoods. Therefore, a hyper-

neighbourhood may have outgoing $\mathcal{C}$-links generated by only one of the "conflicting" relationships. Noticing this countenances the following *exclusion* property.

Lemma 5.4.7 *For any two states $x \in \mathcal{W}$ and $y \in \mathcal{W}$, there is no justifier literal $\overset{\circ}{\epsilon}$ such that both $[\overset{\circ}{\epsilon}] \cap N(x) \subseteq L(x, y)$ and $[\neg\overset{\circ}{\epsilon}] \cap N(x) \subseteq L(x, y)$ hold.*

In this sub-section, we established an important characteristic of the binary relation $\mathcal{C}$ reflected in connection sets L: there is a justifier literal $\overset{\circ}{f}$ such that $[\overset{\circ}{f}] \cap N(x) \subseteq L(x, y)$ for some states x and y, if and only if there exists a causal relationship with literal f being the cause and literal ρ, where $\{\rho\} = y \setminus x$, being the effect. In addition, we verified that if there is a justifier literal $\overset{\circ}{f}$ such that $[\overset{\circ}{f}] \cap N(x) \subseteq L(x, y)$ for some states x and y, then, by the exclusion property, $[\neg\overset{\circ}{f}] \cap N(x) \subseteq L(x, y)$ does not hold.

5.4.2 Propagating in Hyper-space

In this sub-section we mainly investigate the nature of propagation in hyper-state space, where minimal change, triggered by the action, propagates from one hyper- neighbourhood to another. The power-state space propagation is much simpler (power-state to power-state), and will be put to use in the next sub-section.

We previously said that a state x is preferred to a state y in terms of the PMA ordering [70], denoted $x \prec_w y$, if and only if ***Diff***$(x, w) \subseteq$ ***Diff***(y, w), where ***Diff***(p, q) represents the symmetric difference of p and q, i.e., $(p \setminus q) \cup (q \setminus p)$. We also defined the set $min(\prec_w, [E])$ as a subset of post-condition states $[E]$, containing states nearest to the state w in terms of the ordering $\prec_w$. In other words,

$$min(\prec_w, [E]) = \{x \in [E] : \neg\exists y \in [E], y \neq x, y \prec_w x\}.$$

We introduce here one more useful definition. A *trigger set* of hyper-states $s \in \Omega$ is the set where the projection $p(s)$ identifies the nearest states (to the initial state w), among states consistent with the direct effects E, and justifier literals in $h(s)$ capture the initial (immediate) causal context.

Definition 5.4.8 *(Trigger set)*
A trigger set of states $\|E\|_w$ is defined for an initial state $w \in \mathcal{W}$ and an action a, where $\langle C, a, E\rangle$ is an action law, as

$$\{s \in N(q) : q \in \mathcal{W}, q \in min(\prec_w, [E]), h(s) \vdash \overset{\circ}{E}\}$$

where $\prec_w$ is the PMA ordering.

That is, in terms of the PMA ordering, $\|E\|_w$ is the set contained in the hyper- neighbourhood $N(q)$ of state q nearest to the initial state w, and the states $s \in \|E\|_w$ jointly represent the initial (immediate) causal context, i.e., initial causally justified changes triggered by effects E. For instance, if an action law $\langle\{b\}, x, \{\neg b\}\rangle$, is applied to the initial state $\{a, b, c\}$, then the trigger set $\|\{\neg b\}\|_{\{a,b,c\}}$ contains exactly those states enclosed in boxes in Figure 5.5.

The following result is an immediate consequence of this definition.

Lemma 5.4.9 *For any initial state $w \in \mathcal{W}$ and an action a, where $\langle C, a, E\rangle$ is an action law, $\cap_{s\in\|E\|_w} h(s) = \overset{\circ}{E}$.*

In terms of justifier literals, what the states $s \in \|E\|_w$ have in common is precisely the literals in $\overset{\circ}{E}$. In other words, $\|E\|_w = [\overset{\circ}{E}] \cap N(q)$, where $q \in min(\prec_w, [E])$. As was noted, for example, the trigger set $\|\{\neg b\}\|_{\{a,b,c\}}$ contains exactly those states enclosed in boxes in Figure 5.5, represented alternatively as $[\neg\overset{\circ}{b}] \cap N(q)$, where $q = \{a, \neg b, c\} = min(\prec_w, [\neg b])$.

Importantly, any trigger set $\|E\|_w$ is properly contained in the hyper-neighbourhood $N(q)$ of the state q nearest to the initial state among post-condition states $[E]$, that is $\|E\|_w \subset N(q)$.

The intuition behind the trigger set is simple: this set is the starting point of causal propagation. In hyper-state space, the trigger set $\|E\|_w$ contains hyper-states where justifier literals in $\overset{\circ}{E}$ corresponding to the immediate effects E are fixed, and other justifier literals vary. For example, if the action with the post-condition $\neg b$ is executed in the initial state $\{a, b, c\}$, then the state $\{a, \neg b, c\}$ is the nearest to the initial one, and its hyper-neighbourhood (depicted on the left hand-side of Figure 5.5) contains the trigger set $\|\{\neg b\}\|_{\{a,b,c\}}$ — or, in other words, all the hyper-states where the literal $\neg\overset{\circ}{b}$ is fixed, and other justifier literals vary.

As established by Lemma 5.4.9, if one takes the intersection of hyper-parts of all the hyper-states in the trigger set, what is left is precisely the literals $\overset{\circ}{E}$ corresponding to

the immediate effects E — in the example, $\overset{\circ}{E} = \{\neg\overset{\circ}{b}\}$, where the literal $\neg\overset{\circ}{b}$ is common to all hyper-states in the trigger set (the states enclosed in boxes in Figure 5.5). Not surprisingly, we wish to start our propagation from the states bounded by the trigger set (where only immediate causes and their justifier literals are fixed), and not from some hyper-states that are not consistent with literals in $\overset{\circ}{E}$.

Moreover, in power-state space we start propagation from the power-state $\gamma(\|E\|_w)$ corresponding to the trigger set. For example, the power-state corresponding to the set $\|\{\neg b\}\|_{\{a,b,c\}}$ is the state mapped from the partial hyper-state $\{a, \neg b, c, \neg\overset{\circ}{b}\}$. In other words, we wish to start our propagation from the power-state reflecting only immediate causes ($\neg b$) and their justifier literals ($\neg\overset{\circ}{b}$).

With the trigger set $\|E\|_w$ defined, we can formally trace a causal propagation in the hyper-state space Ω.

Definition 5.4.10 *(Causally triggered hyper-neighbourhood)*
We say that a hyper-neighbourhood $N(q)$, where $q \in \mathcal{W}$, is causally triggered by the set $\|E\|_w$, denoted as $\|E\|_w \succ N(q)$ if and only if for all $s \in \|E\|_w$, there exists $s' \in N(q)$, such that $C^*(s, s')$ holds.

Basically, a hyper-neighbourhood $N(y)$ is causally triggered when *all* initially justified causal changes propagate to a subset of $N(y)$.

Returning to our example of Figure 5.5 we see that (assuming a direct action effect $\neg b$) this was an instance when the trigger set $\|\{\neg b\}\|_{\{a,b,c\}}$ triggers the hyper- neighbourhood on the right-hand side.

Figure 5.6 gives an example when the same trigger set fails to trigger the same hyper-neighbourhood. This is due to the fact that, in this case, not all states in $\|\{\neg b\}\|_{\{a,b,c\}}$ belong to the connection set. We can easily verify that the causal relationship used to generate the connection set in this example would not be applicable according to Thielscher's approach also. This is because the cause (c) is not a part of the history component which is equal to the direct effect ($\neg b$) at this stage.

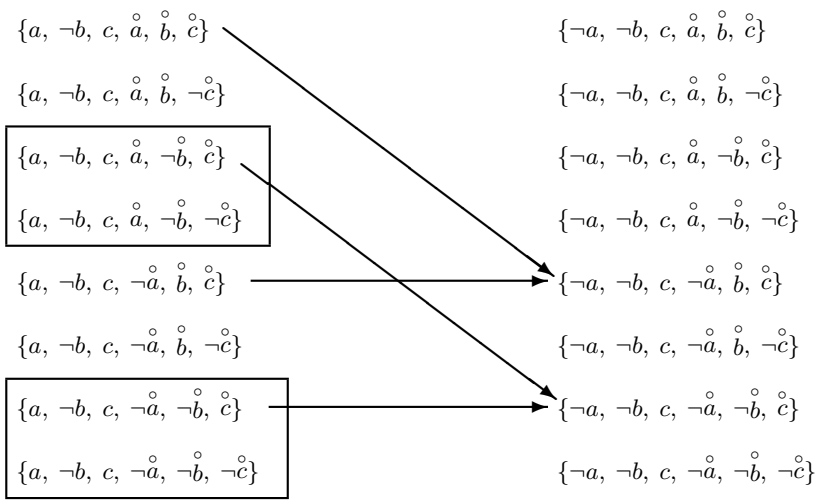

Figure 5.6: The $\mathcal{C}$-links between hyper-neighbourhoods of the states $\{a, \neg b, c\}$ and $\{\neg a, \neg b, c\}$, generated by a causal relationship c *causes* $\neg a$ *if* $\neg b$. Some $[\neg\mathring{b}]$-states do not belong to the connection set $L(\{a, \neg b, c\}, \{\neg a, \neg b, c\})$.

5.4.3 Locating Successor States

We can view changes triggered by the set $\|E\|_w$ as propagating in hyper-state space towards a hyper-neighbourhood of a possible successor state, passing through some (causally triggered) hyper-neighbourhoods on the way. The point where this propagation ends can now be defined explicitly.

Definition 5.4.11 *(Final state)*
A hyper-state $s \in \Omega$ is final *if and only if* $\{s' : \mathcal{C}(s, s')\} = \emptyset$.
A set $z \subseteq N(r)$ for some $r \in \mathcal{W}$ is final *if and only if some hyper-state $s \in z$ is final.*
A power-state $x \in \Gamma$ is final *if and only if* $\{x' : x \rightharpoonup x')\} = \emptyset$.

Importantly, in defining final *sets* of hyper-states we did not require that all elements are final, but rather, that only some (at least one) are final.

Let us revisit our original example with the causal relationship $\neg b$ *causes* $\neg a$ *if* $\top$, and extend it with another causal relationship c *causes* b *if* $\neg a$. The case is depicted in Figure 5.7, where the $\mathcal{C}$-links out-going from the right-hand side hyper-neighbourhood

are generated by the added causal relationship (these links connect to some other appropriate hyper-states omitted from the figure). Again, let us consider the initial state $w = \{a, b, c\}$, and the action with the direct effect $\neg b$. The trigger set $\|\{\neg b\}\|_w$ triggers the hyper-neighbourhood $N(q')$ of the state $q' = \{\neg a, \neg b, c\}$ on the right-hand side. The hyper-states enclosed in a box belong to the transitively traced set $T^*(\|\{\neg b\}\|_w, q')$. The second causal relationship is not applicable because c is not a part of the current effects (history) $\{\neg b, \neg a\}$. The hyper-state space equivalent of this fact is the absence of an out-going $\mathcal{C}$-link from at least one state in the set $T^*(\|\{\neg b\}\|_w, q')$. This makes the latter set final.

The hyper-states in the set $T^*(\|\{\neg b\}\|_w, q')$ agree on justifier literals $\neg\overset{\circ}{a}$ and $\neg\overset{\circ}{b}$, while the value of the literal $\overset{\circ}{c}$ varies. If there were out-going $\mathcal{C}$-links from *all* hyper-states in the transitively traced set, the latter would not be final. Let us demonstrate that in this case there would have to be another (third) *applicable* causal relationship, continuing propagation further in Thielscher's approach as well.

The out-going $\mathcal{C}$-links from both hyper-states in the transitively traced set would be generated by some other relationship. This relationship could not have $\neg c$ as its cause literal simply because the proper part $p(q') = \{\neg a, \neg b, c\}$ is inconsistent with $\neg c$ — this is, in fact, a key point of the exclusion property formalised by Lemma 5.4.7. If, on the other hand, this third relationship had c as its cause literal, it could not generate any $\mathcal{C}$-link from a hyper-state containing $\neg\overset{\circ}{c}$.

Hence, this relationship must have either $\neg a$ or $\neg b$ as its cause literal. In either case it becomes applicable, because both $\neg a$ and $\neg b$ belong to the current effects (history). Therefore, propagation would continue according to Thielscher's approach too.

Having described the intuition behind final points of propagation, we are now in a position to define the set of successor states $Res_\Omega(w, a)$ according to the hyper-state space semantics. We shall see that this completely characterises Thielscher's resultant state set $Res_{RDL}(w, a)$.

Definition 5.4.12 ($Res_\Omega(w, a)$)
Let $\mathcal{F}$, A, D, R, $\mathcal{L}$, w, $\langle C, a, E\rangle$ be the same as in Definition 5.1.3, Ω the set of hyper-states, and $\mathcal{C}$ a causal binary relation defined by Definition 5.2.6. A state $r \in \mathcal{W}$,

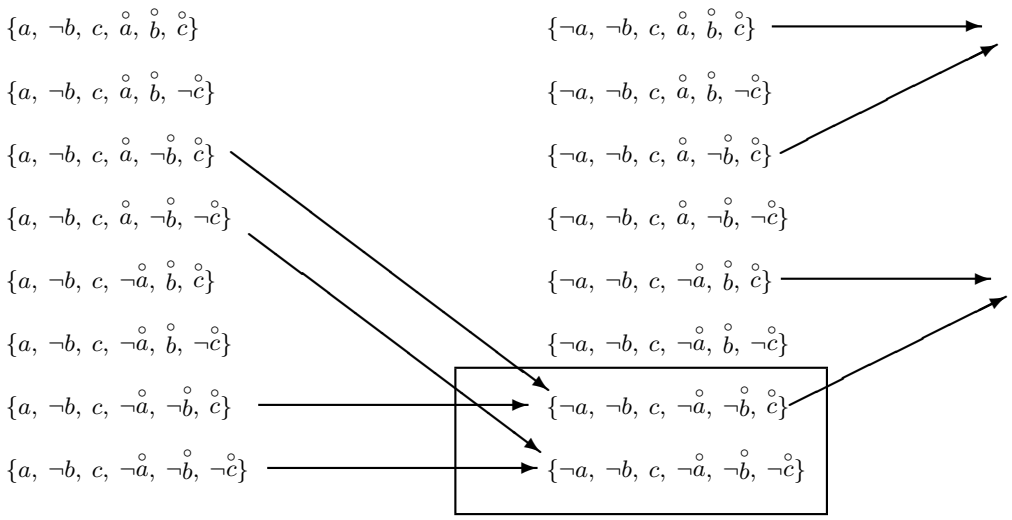

Figure 5.7: The final transitively traced set.

satisfying D, is a successor state *of w and a, denoted $r \in Res_\Omega(w, a)$, if and only if there exists an applicable (with respect to w) action law $\langle C, a, E\rangle$ such that $\|E\|_w \succ N(r)$ and $T^*(\|E\|_w, r)$ is final.*

Equivalently,

$$Res_\Omega(w, a) \equiv \{r \in \mathcal{W} : r \text{ satisfies } D,\ \|E\|_w \succ N(r),\ T^*(\|E\|_w, r) \text{ is final }\}.$$

In other words, a hyper-neighbourhood of a successor state must be causally triggered by the trigger set, and the transitively traced set of the latter must be final (i.e., it must contain at least one final hyper-state).

The following lemma will prove useful in arriving at the representation results we are seeking.

Lemma 5.4.13 *If $\|E\|_w \subset N(x)$, then $\|E\|_w \succ N(y)$ for some $y \in \mathcal{W}$ if and only if $(x, E) \overset{*}{\rightsquigarrow} (y, E')$ for some E'.*

This lemma provides an important link between Thielscher's approach and propagation in the hyper-state space, pointing out that causal propagation in normal state-space (complemented by an explicit account of all changes) is equivalent to propagation along causally-triggered hyper-neighbourhoods. We can now state one of the results of central importance in this chapter.

Theorem 5.4.14 $Res_{RD\mathcal{L}}(w, a) = Res_{\Omega}(w, a)$.

There is an interesting and useful dependency between propagation towards causally triggered hyper-neighbourhoods (in the hyper-state space) and propagation towards power-states corresponding (in the power-state space) to transitively traced sets. This dependency is given by the following corollary.

Corollary 5.4.15 *For any states* $w \in \mathcal{W}$ *and* $q \in \mathcal{W}$, $\|E\|_w \succ N(q)$ *if and only if* $\gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(T^*(\|E\|_w, q))$.

In other words, whenever a hyper-neighbourhood $N(q)$ is causally triggered by the trigger set, the power-state corresponding to the transitively traced set $T^*(\|E\|_w, q)$ is "reachable" by means of the binary relation $\rightharpoonup$ from the power-state corresponding to the trigger set.

We would like to conclude with a representation result for the power-state space semantics. The following definition establishes the selection function for this semantics.

Definition 5.4.16 ($Res_{\Gamma}(w, a)$)
Let $\mathcal{F}$, A, D, R, $\mathcal{L}$, w, $\langle C, a, E\rangle$ *be the same as in Definition 5.1.3,* Γ *the set of power-states, and* $\rightharpoonup$ *be binary relation defined by Definition 5.3.2. A state* $r \in \mathcal{W}$*, satisfying* D*, is a* successor state *of* w *and* a*, denoted* $r \in Res_{\Gamma}(w, a)$*, if and only if* $\gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(z)$*, where* $z \subseteq N(r)$ *and* $\gamma(z)$ *is final.*

Equivalently,

$$Res_{\Gamma}(w, a) \equiv \{r \in \mathcal{W} : r \text{ satisfies } D,\ \gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(z),\ z \subseteq N(r),\ \gamma(z) \text{ is final }\}.$$

Intuitively, the causal propagation in power-state space starts in the power-state which corresponds to the trigger set of hyper-states (analogous to the gradient area in the augmented preferential semantics), and ends in the final power-state corresponding to the transitively traced set of *all* initial justifier literals (or any subset of this transitively traced set). In other words, this process simply propagates "minimal change" within the space of possible power-states, instead of keeping an explicit (and changing) account of context-dependent effects of actions.

It is important to realise that, although the underlying construction and proofs are fairly complex, and the dimension of the power-state space Γ is large, the semantics remains simple. It describes, in simple terms, a process of propagation from the minimal elements (gradient area) to final state(s), followed by projection onto the standard space $\mathcal{W}$.

The second central result of this chapter can now be obtained.

Theorem 5.4.17 $Res_\Omega(w,a) = Res_\Gamma(w,a)$.

Analogous results can be obtained for conservative successor states as well. Defining

$$Res^*_\Omega(w,a) = \{r \in Res_\Omega(w,a),\ E \subseteq r\}$$
and
$$Res^*_\Gamma(w,a) = \{r \in Res_\Gamma(w,a),\ E \subseteq r\},$$

we obtain:

Theorem 5.4.18

$Res^*_{RD\mathcal{L}}(w,a) = Res^*_\Omega(w,a)$.
$Res^*_\Omega(w,a) = Res^*_\Gamma(w,a)$.

It is quite clear that the power-space semantics is an instance of the augmented preferential semantics. The precise reduction will be described in Chapter 7, by equating the information state-space with the power-state space Γ, setting the binary causal relation $\mathcal{M} = \rightharpoonup$, and defining an appropriate projection function $\mathcal{P}$ and preferential structure $\mathcal{O}$.

5.5 Discussion

We have described here a novel type of semantics for a particular causal approach to reasoning about action. The basic idea is to abandon the standard state-space of possible worlds and consider instead a larger set of possibilities — a power-state space — tracing

the effects of actions (including indirect effects) with the states in the power-state space. These additional states are reminiscent of a semantics (that included a labelling function) provided by Kraus et al. [24] for non-monotonic consequence relations. Intuitively, the purpose of these power-states is to supply extra context to record the process of causal propagation. More precisely, we propose to use a binary (causal) relation on states, while propagating "minimal change" within a set of possible states, instead of keeping an explicit (and changing) account of context-dependent effects of actions. Essentially, the "minimal change" (encapsulated in the states consistent with the direct effects of an action) triggers the process of causal propagation which continues towards a final state.

The hyper-state space semantics gives the original motivation behind the approach adopted here and allowed us to establish the initial results. It enables simple causal propagation through states of higher dimension, encoding additional contextual information. This is achieved through the introduction of extra fluents; one for each original fluent in the underlying object language. Abstracting away the important features of the hyper-state space semantics led us to the power-state space semantics. The latter gives a more direct and arguably natural view of causal propagation, being a variant of the augmented preferential semantics. Equipped with the proposed semantics, we believe that we can advance toward a unifying framework for our motivating approaches.

Chapter 6

Causal Propagation Semantics

We have shown in the previous chapters that a pure preferential semantics alone was not capable of characterising two influential approaches to reasoning about action and causality in a uniform way. On the other hand, variants of preferential semantics augmented by additional structures on the state space have been successfully used to capture the considered approaches. To re-iterate, McCain and Turner's causal theory of action [37] was characterised by a variant of the augmented preferential semantics, using an appropriately constructed binary relation on states in addition to a preference relation. This additional relation captured causal context in action systems by translating individual causal laws (rules) into causality-driven state transitions. Subsequently, Thielscher's causal theory of action [63] has been characterised by another variant of the augmented preferential semantics. This time the minimality component was complemented by a binary relation on states of higher dimension. The effects of actions (including indirect ones) were traced in the information (power-state) space. Again, the purpose of power-states was to supply extra context for the process of causal propagation. The furnished power-state space semantics can be clearly seen to employ a component of minimal change coupled with causality.

There is, however, another prominent framework — Sandewall's causal propagation semantics [56] — that has attempted to provide a unifying framework to theories of action and causality. At a first glance, the causal propagation semantics is not a preferential style semantics — it does not contain an explicit component ordering possible states of the world, or any other preferential structure, for that matter.

The primary aim of this chapter is to uncover the Principle of Minimal Change hidden, we believe, behind action invocation and causal propagation in Sandewall's framework. In other words, by studying the causal propagation semantics proposed by Sandewall [56], we intend to demonstrate that minimal change and causality co-exist in separate roles and enhance each other in Sandewall's framework as well.

This will further our understanding of causality as introduced by various proposals and advance us towards a generalised semantics.

6.1 Technical Background

The assessment of ramification methods that use static domain constraints [56] demonstrated a need for an underlying semantics. The suggested approach introduced a *causal propagation semantics* as a necessary prerequisite for the analysis of the sound applicability of various propagation-oriented ramification methods. The purpose of this semantics was to define a set of intended models in "a precise, simple and intuitively convincing fashion" [56]. In other words, this semantics advanced us towards a concise solution to the frame and ramification problems.

In this section we reproduce, for convenience, the technical preliminaries described in Chapter 2.

Sandewall uses the following basic concepts. The set of possible states of the world, formed as a Cartesian product of the finite sets of a finite number of state variables, is denoted as $\mathcal{W}$. $\mathcal{E}$ is the set of possible actions. The causal propagation semantics extends a basic state transition semantics with a *causal transition relation*. The causal transition relation C is a non-reflexive relation on states in $\mathcal{W}$. A state r is called *stable* if it does not have any successor s such that $C(r, s)$; we will denote the set of stable states $\{r \in \mathcal{W} : \neg\exists s \in \mathcal{W}, C(r, s)\}$ as $\mathcal{S}_c$. Another component, $\mathcal{D}$, is the set of admitted states chosen as a subset of $\mathcal{S}_c$. The set $\mathcal{D}$ may or may not be chosen to contain all the stable states—in other words, some stable states may not be admitted.

Another important concept, introduced by Sandewall, is an *action invocation relation* $G(e, r, r')$, where $e \in \mathcal{E}$ is an action, r is the state where the action e is invoked, and r' is "the new state where the instrumental part of the action has been executed" [56]. In

other words, the state r' satisfies the direct effects of the action e. It is required that every action is always invokable, that is, for every $e \in \mathcal{E}$ and $r \in \mathcal{W}$ there must be at least one r' such that $G(e, r, r')$ holds. Of course, this requirement does not mean to guarantee that every action results in an admitted state—on the contrary, the intention is to trace the indirect effects of the action, potentially leading to an admitted (and, therefore, stable) state.

A finite (the infinite case is omitted) transition chain for a state $w \in \mathcal{D}$ and an action $e \in \mathcal{E}$ is a finite sequence of states $r_1, r_2, ..., (r_k)$, where $G(e, w, r_1)$ and $C(r_i, r_{i+1})$ for every $i, 1 \leq i < k$, and where r_k is a stable state. The last element of a finite transition chain is called a result state of action e performed in state w. We assume here that C is a non-empty relation — otherwise a transition chain is (technically) undefined (there must be at least one propagation step, according to the given definition) and no result states are possible[1].

These basic concepts define an *action system* as a tuple $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G \rangle$. As with many other state transition action systems, the intention is to characterise a result state r_k in terms of the initial state r and action e, without "referring explicitly to the details of the intermediate states" [56]. In other words, it is desirable to define a selection function $Res(r, a)$. For an action system $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G \rangle$, a function selecting states resulting from the action e performed at the state $w \in \mathcal{D}$, can be given as

$$Res_{CG}(w, e) = \{r_k \in \mathcal{S}_c : G(e, w, r_1) \text{ and } C(r_i, r_{i+1}), 1 \leq i < k\}.$$

The following definition strengthens action systems based on the causal propagation semantics.

Definition 6.1.1 *($\lhd_w(p, q)$)*
If three states w, p, q are given, we say that the pair p, q respects w, denoted as $\lhd_w(p, q)$, if and only if $p(f) \neq q(f)$ implies $p(f) = w(f)$ for every state variable f that is defined in $\mathcal{W}$, where $r(f)$ is a valuation of variable f in state r.

[1] Alternatively, one can extend the definition of a transition chain to cover the simplest case — when the state r_1 in $G(e, r, r_1)$ is stable.

Definition 6.1.2 *(Respectful action system)*
An action system $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G \rangle$ is called respectful if and only if, for every $w \in \mathcal{D}$, every $e \in \mathcal{E}$, w is respected by every pair r_i, r_{i+1} in every transition chain, and the last element of the chain is a member of $\mathcal{D}$.

According to Sandewall [56], respectful action systems are intended to ensure that in each transition there cannot be changes in state variables which have changed previously upon invocation or in the causal propagation sequence. The requirement of "respectfulness" will be analysed in detail, in Section 6.4.

We mentioned earlier that, while there are certain similarities between the causal propagation semantics and our augmented preferential semantics (including components such as $\mathcal{W}$, $\mathcal{E}$, $\mathcal{D}$, and causal transition relation), it is not obvious that they define the same successor states. In particular, the Principle of Minimal Change is not explicit in the causal propagation semantics.

6.2 Invoking Minimal Change

Our primary focus will be discovering and capturing the nature of minimality hidden, as we believe, in the invocation relation G.

Motivated by a preferential-style semantics, one may be tempted to suggest a set of orderings $<_w$, each with respect to some $w \in \mathcal{W}$, such that the invocation relation G for a given state w can be simply realised by selecting the nearest state(s) to w satisfying an action post-condition (the state where the instrumental part of the action has been executed). Again, let $[e]$ denote a set of states satisfying the post-conditions of an action e. A set $min(<_w, [e])$ is defined as a subset of $[e]$ containing states nearest to the set w in terms of the ordering $<_w$. In other words,

$$min(<_w, [e]) = \{p \in [e] : \neg \exists q \in [e], q \neq p, q <_w p\}.$$

As usual, we will refer to an element of $min(<_w, e)$ as a $<_w$-minimal state in $[e]$.

However, it does not appear possible to realise the invocation relation G for a given state w through selecting the nearest state(s) to w among states in $[e]$. More precisely, the following observation can be obtained.

Lemma 6.2.1 *There is no ordering $<_w$ such that for every action e and state r, $G(e, w, r)$ if and only if $r \in min(<_w, [e])$.*

Proof:

Assume that there exists an ordering $<_w$ such that for every action e and state r, $G(e, w, r)$ if and only if $r \in min(<_w, [e])$, that is $\neg\exists x \in [e], x \neq r$, such that $x <_w r$. Let w, s, p, q be states of an action system, and e_1, e_2 be its actions. Let $G(e_1, s, p)$, $G(e_1, s, q)$, $G(e_1, w, p)$, $G(e_1, p, q)$, $G(e_1, q, p)$ be the only invocations of the action e_1. These invocations indicate that the states p and q are the only states satisfying the post-conditions of the action e_1, in other words, $\{p, q\} = [e_1]$. But when the action is executed at the state w, only the state p is selected. Hence, $\neg(q <_w p)$. Analogously, let $G(e_2, s, p)$, $G(e_2, s, q)$, $G(e_2, w, q)$, $G(e_2, p, q)$, $G(e_2, q, p)$ be the only invocations of the action e_2, indicating that $\{p, q\} = [e_2]$. A similar argument leads to the conclusion $\neg(p <_w q)$. Since $[e_1] = [e_2] = \{p, q\}$, it follows that $G(e_1, w, q)$ and $G(e_2, w, p)$ should also be included among invocations of this action system, which is not the case.

■

This means that it is not possible to define the relation G in terms of a preference relation on states without imposing some restrictions.

Alternatively, action post-conditions must be conditionalised on states of invocation. The last suggestion follows the action languages approach [23], where an action is defined in an effect proposition

$$\text{A } \textbf{causes } \varphi \textbf{ if } \psi,$$

where A is an action, and φ and ψ are fluent formulae; post- and pre-conditions respectively. We did not wish to adopt the latter tactic, as it would weaken our quest for conciseness in seeking a solution to the frame problem. Consequently, our intention at this stage is to restrict the invocation relation G in such a way that, given an initial state and an action, the invoked states can be characterised precisely as states nearest to the initial one in terms of some appropriate minimality ordering. This characterisation will highlight the role of minimal change in the causal propagation semantics under consideration, and open the way to uniformly compare this semantics with our motivating approaches.

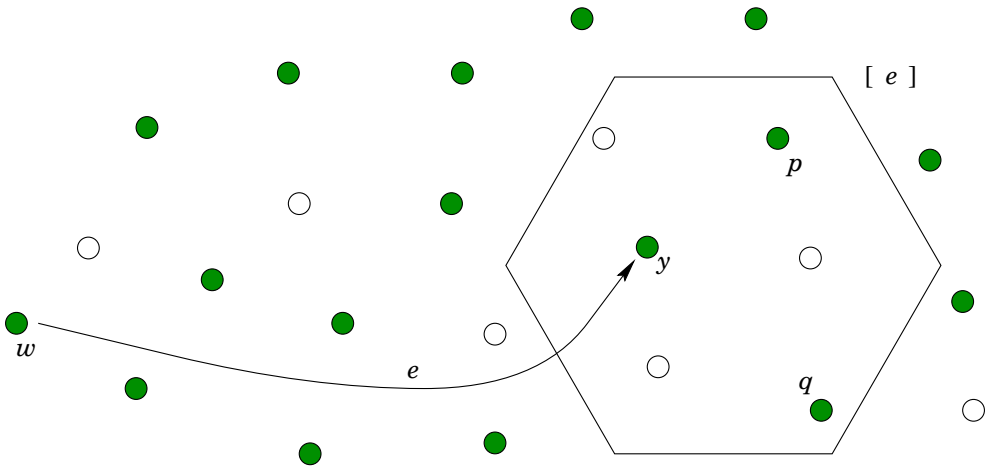

Figure 6.1: The action e is a $\{y, p, q\}$-covering action, and the state y is a $\{y, p, q\}$-cover accessible state, given $G(e, w, y)$ (shown as the curved arrow marked with e).

Before we identify the required restrictions on the invocation relation G, we introduce some more abbreviations.

Definition 6.2.2 *(S-covering action)*
For a set of states $S \subseteq \mathcal{W}$ and an action $a \in \mathcal{E}$, the action a is called an S-covering *action if and only if* $S \subseteq [a]$.

Definition 6.2.3 *(S-cover accessible state)*
For states $w, x \in \mathcal{W}$, we say that the state x is S-cover accessible *from state w, if and only if there exists an S-covering action a such that $G(a, w, x)$.*

Furthermore, we say that a state x is *not S-cover accessible* from state w, if there is no S-covering actions a such that $G(a, w, x)$.

Importantly, it follows that all states in a set S satisfy the post-conditions of an S-covering action. It is worth pointing out that, given two states p and q satisfying the post-conditions of some action a (that is, the action a is a $\{p, q\}$-covering action), the state p may be $\{p, q\}$-cover accessible from some state w, while state q is not $\{p, q\}$-cover accessible from w. This is the case when the following two conditions are satisfied: $\exists a \in \mathcal{E}, p, q \in [a], G(a, w, p)$ and $\forall e \in \mathcal{E}, p, q \in [e], \neg G(e, w, q)$.

The first restriction on the invocation relation is given as

(G_1) if p is $\{p, q\}$-cover accessible from w but q is not, and

q is $\{q, x\}$-cover accessible from w but x is not

then p is $\{p, x\}$-cover accessible from w and x is not,

for arbitrary states w, p, q, x.

The premise of the implication is that, considering all actions whose post-conditions are satisfied by two states p and q, state p is chosen at least once by the invocation relation and state q is never chosen; and considering all actions whose post-conditions are satisfied by two states q and x, state q is chosen at least once, while state x is never chosen. This then necessitates that, considering all actions whose post-conditions are satisfied by states p and x, invocation of the state p must eventuate at least once, but state x cannot be invoked at all.

At this stage, it might be useful, albeit syntactically cumbersome, to express the condition (G_1) purely in terms of the invocation relation G.

$$\forall\, w, p, q, x \in \mathcal{W},$$

$$\begin{gathered}(\exists a \in \mathcal{E}, p, q \in [a], G(a, w, p)) \wedge (\forall e \in \mathcal{E}, p, q \in [e], \neg G(e, w, q)) \\ \wedge\, (\exists a' \in \mathcal{E}, q, x \in [a'], G(a', w, q)) \wedge (\forall e' \in \mathcal{E}, q, x \in [e'], \neg G(e', w, x)) \\ \supset \\ (\exists a'' \in \mathcal{E}, p, x \in [a''], G(a'', w, p)) \wedge (\forall e'' \in \mathcal{E}, p, x \in [e''], \neg G(e'', w, x))\end{gathered}$$

Undoubtedly and not surprisingly, the condition (G_1) has a transitive flavour. It becomes even clearer if one uses the requirement that, considering all actions whose post-conditions are satisfied by two states p and q, state p is chosen at least once by the invocation relation and state q is never chosen (that is, p *is* $\{p, q\}$-cover accessible from some w, but q is *not*), as a preference criterion. This will be formally captured later.

The condition (G_1) does not, however, rule out the counter-example used in the proof of Lemma 6.2.1. Another condition is needed and is given as

(G_2) Given any two $\{p, q\}$-covering actions e' and e'',

if $G(e', w, p)$ and $G(e'', w, q)$ then $G(e', w, q)$.

This condition simply requires that if neither of two states p and q is chosen over the other in terms of the criterion implicitly used in the condition (G_1), then selection of either of them necessitates selection of the other. It is interesting to contrast this condition with the "irrelevance of syntax" condition which requires, for two actions e' and e'' such that $[e'] = [e'']$ and any states w and q, that if $G(e', w, q)$ then $G(e'', w, q)$. In other words, the irrelevance of syntax condition postulates that if two actions agree in terms of the post-condition states, then they have to agree in terms of invocation states. Clearly, if the irrelevance of syntax condition holds then the condition (G_2) holds as well — simply because any two actions agreeing on post-conditions would "cover" the same sets of states. However, the condition (G_2) does not necessarily enforce the irrelevance of syntax condition, making the former a weaker manifestation of the latter.

It is easy to verify that (G_2) is logically equivalent to

$$\forall e \in \mathcal{E}, p, q, w \in \mathcal{W}, p, q \in [e],$$

$$\neg(((\forall e' \in \mathcal{E}, p, q \in [e'], \neg G(e', w, q)) \wedge (\exists a \in \mathcal{E}, p, q \in [a], G(a, w, p)))$$
$$\vee$$
$$((\forall e' \in \mathcal{E}, p, q \in [e'], \neg G(e', w, p)) \wedge (\exists a' \in \mathcal{E}, p, q \in [a'], G(a, w, q))))$$
$$\wedge G(e, w, p) \supset G(e, w, q)$$

Here, the criterion

$$\forall e' \in \mathcal{E}, p, q \in [e'], \neg G(e', w, q)) \wedge (\exists a \in \mathcal{E}, p, q \in [a], G(a, w, p),$$

favouring state p over state q is made explicit, and the negation on the left-hand side specifies that neither of states p and q is chosen over the other.

Clearly, the condition (G_2) rules out the counter-example, used in the proof of Lemma 6.2.1. Since the example postulated $G(e_1, w, p)$, $G(e_2, w, q)$, and $p, q \in [e_1] \cap [e_2]$, the condition (G_2) therefore would imply $G(e_1, w, q)$ and $G(e_2, w, p)$ which are not present in the action system.

Another useful (contrapositive) form of the condition (G_2) is given as

(G_2') Given any two $\{p, q\}$-covering actions e' and e'',

if $G(e', w, p)$ and not $G(e', w, q)$ then not $G(e'', w, q)$.

In other words, if the invocation of the action e' at the state w leads to one state p and not to another state q, then the invocation of the action e'' (that agrees with the action e' in terms of covering the set $\{p, q\}$) cannot lead to the state q as well.

Finally, we reinforce Sandewall's constraint that any action is invokable in principle.

(G_3) $\forall\, e \in \mathcal{E},\ w \in \mathcal{W},\ \exists\, p \in [e],\ G(e, w, p)$

As noted above, this condition does not guarantee that the invoked action will succeed — it may possibly be qualified by causal propagation ending in a non-admitted state.

In summary, we presented here three intuitive uniformity principles restricting the invocation relation G. The aim of these restrictions is to ensure that instead of the invocation relation one may simply use some preferential structure selecting the same states, for each initial state. Of course, the corresponding preferential relations would have to satisfy certain condition as well. Now, we are in a position to describe a set of orderings $\mathcal{O}$ corresponding to the invocation relation. Ideally, any corresponding ordering $<_w$ should satisfy only the transitivity property:

(M_1) if $p <_w q$ and $q <_w x$ then $p <_w x$.

However, it turns out that, given an action system, the related ordering has to satisfy, in addition, two other properties.

(M_2) if p is $<_w$-minimal in $[a]$ for some $\{p, q\}$-covering action a and

q is not $<_w$-minimal in $[e]$ for any $\{p, q\}$-covering action e

then $p <_w q$.

(M_3) if $p <_w q$

then p is $<_w$-minimal in $[e]$ for some $\{p, q\}$-covering action e.

Basically, the second property (M_2) requires that any state p which is $<_w$-minimal in some set $[a]$ is preferred to any state q, where q belongs to the set $[a]$ as well, and which is not $<_w$-minimal in any set $[e]$ of action post-conditions satisfied by both p and q.

The third property (M_3) posits that if a state p is preferred to a state q by a preference relation $<_w$, then there must exist an action e, whose post-conditions are satisfied by these two states, such that state p is $<_w$-minimal in $[e]$.

Expanded equivalents of conditions (M_2) and (M_3) are given below:

$$(M_2) \quad \forall w, p, q \in \mathcal{W}, \quad (\exists a \in \mathcal{E}, p, q \in [a], p \in min(<_w, a)) \\ \wedge \ (\forall e \in \mathcal{E}, p, q \in [e], q \notin min(<_w, e)) \ \supset \ p <_w q$$

(M_3) $\quad \forall w, p, q \in \mathcal{W}, \quad p <_w q \ \supset \ \exists e \in \mathcal{E}, p, q \in [e]$, such that $p \in min(<_w, e)$

These conditions may seem to be quite restrictive. For example, the PMA ordering [70] does not satisfy condition (M_2): a minimal state may not necessarily be preferred to a non-minimal one. However, this condition is needed if we want to define the invocation relation in terms of a preference relation in a straightforward manner, as will be illustrated later.

We intend to prove at this stage that there is a way to define the invocation relation in terms of a preference relation and vice versa, while preserving the respective selections of states satisfying the direct effects of an action. The following two definitions will be shown to ensure such an equivalence.

Definition 6.2.4 *A new invocation relation $G_<$ is defined as follows: $G_<(e, w, r)$ if and only if r is $<_w$-minimal in $[e]$, where $w, r \in \mathcal{W}, e \in \mathcal{E}$.*

Put simply, the new relation $G_<(e, w, r)$ specifies states r that are nearest among all states in $[e]$ to the initial state w, where the action e was invoked.

Definition 6.2.5 *Given an invocation relation G, for each $w \in \mathcal{W}$ we define an ordering $<_{w,G}$ on states in $\mathcal{W}$ as follows: $p <_{w,G} q$ if and only if state p is $\{p, q\}$-cover accessible from w and state q is not $\{p, q\}$-cover accessible from w.*

This definition specifies a preference relation on states generated by a given invocation relation—state p is nearer to an initial state w than state q if and only if for all actions whose direct effects are satisfied by both states p and q, the state q is never selected by the invocation relation G, while state p is selected at least once.

6.3 Representation Results

The following two lemmas establish the sought-after equivalence between invocation and preference relations.

Lemma 6.3.1 *If the relation G satisfies the conditions $(G_1) - (G_3)$, then for each $w \in \mathcal{W}$, the ordering $<_{w,G}$ satisfies conditions $(M_1) - (M_3)$.*

Lemma 6.3.2 *If each ordering $<_w$ for $w \in \mathcal{W}$ satisfies conditions $(M_1) - (M_3)$, then the relation $G_<$ satisfies the conditions $(G_1) - (G_3)$.*

Lemma 6.3.1 provides support for a translation from an action system, based on Sandewall's approach and relying on the invocation relation G in order to produce states consistent with the direct effects of an executed action e, to an action system which uses for this purpose a newly constructed preference relation $<_{w,G}$, selecting states that are minimal among states satisfying the actions post-conditions.

The parallel Lemma 6.3.2 underlies a reverse translation where given orderings $<_w$ produce an invocation relation $G_<$, while preserving selection of states minimal in terms of $<_w$.

Since the relation $G_<$ is constructed from the set of orderings $<_w$ for each $w \in \mathcal{W}$, and the orderings $<_{w,G}$ are constructed for each $w \in \mathcal{W}$ from the invocation relation G, we can in turn construct, for each $w \in \mathcal{W}$, another ordering $<_{w,G_<}$ from the relation $G_<$, and another invocation relation $G_{<_{w,G}}$ from the set of orderings $<_{w,G}$. As expected, the following identity properties can be established.

Lemma 6.3.3 $G_{<_{w,G}}(e, w, r)$ *if and only if* $G(e, w, r)$.

Lemma 6.3.4 *For each ordering $<_w$,* $p <_{w,G_<} q$ *if and only if* $p <_w q$.

Lemma 6.3.3 maintains that a given invocation relation G is preserved by a process, which uses it in producing a preference relation $<_{w,G}$ by Definition 6.2.5, with a subsequent construction of a relation $G_{<_{w,G}}$ by Definition 6.2.4.

The identity in Lemma 6.3.4 indicates that a process, starting with orderings $<_w$ for each $w \in \mathcal{W}$, transformed into an invocation relation $G_<$ by Definition 6.2.4, with a subsequent construction of preference relations $<_{w,G_<}$ by Definition 6.2.5, preserves an original ordering $<_w$ for each $w \in \mathcal{W}$.

6.4 Causal Propagation in Respectful Action Systems

Having "discovered" the role of minimality in the process of action invocation, we now proceed to an analysis of actual propagation in the state-space, driven by causal chains after an action is invoked.

According to Sandewall, respectful action systems are intended to work as follows [56]:

> Suppose the world is in a stable state r, and an action E is invoked. The immediate effect of this is to set the world in a new state, which is not necessarily stable. If it is not, then one allows the world to go through the necessary sequence of state transitions, until it reaches a stable state. *That whole sequence* of state transitions is together viewed as the action, and the resulting admitted state is viewed as the result state of the action.

In particular, what a respectful action system tries to ensure is that in each transition there cannot be changes in state variables which have changed previously upon invocation or in the causal propagation sequence. This requirement, of course, guarantees that a resultant state is always consistent with the direct effects of the action (which cannot be cancelled by indirect ones), and that there are no cycles in transition chains.

It is important to analyse at this stage, how much the "respectfulness" requirement restricts an action system $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G \rangle$.

First of all, it is interesting to observe that the respectfulness requirement in terms of states is related to the notion of minimality as well. More precisely, the former can be

achieved by a preference relation on states. Let us recall that a state x is preferred to a state y in terms of the PMA ordering [70], denoted $x \prec_w y$, if and only if *Diff*$(x, w) \subset$ *Diff*(y, w), where *Diff*(p, q) represents the symmetric difference of p and q, i.e., $(p \setminus q) \cup (q \setminus p)$. Formally, the following observation holds.

Lemma 6.4.1 *The pair p, q respects w if and only if $p \prec_w q$ in the PMA ordering $\prec_w$ associated with w.*

Let us denote the fact that a pair p, q respects a state w by $\lhd_w(p, q)$. This lemma established that $\lhd_w(p, q)$ if and only if $p \prec_w q$.

As defined in Section 6.1 (Definition 6.1.1), an action system $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G \rangle$ is said to be respectful if and only if, for every $w \in \mathcal{D}$, every $e \in \mathcal{E}$, w is respected by every pair q, s in every transition chain, and the last element of any finite chain is a member of $\mathcal{D}$.

Lemma 6.4.1 indicates that in a respectful action system causality propagates from closer states to states which are more distant from the initial one, in terms of the PMA ordering. This explains why a resultant state always satisfies the direct effects of an action: any reversal in truth value of a changed state variable would mean a propagation backward to a PMA-closer state, which is ruled out in a respectful action system. Consequently, direct effects stay unchanged in any transition chain as they have been changed once on invocation.

The given definition defines a strongly respectful system — when an admitted state has to be respected by any pair in *every* transition chain. We believe, however, that Sandewall is likely to have intended that an admitted state w is required to be respected by every pair in every transition chain *for the state* w, and not in every other transition chain. This interpretation defines a *weakly respectful* action system.

Definition 6.4.2 *(Weakly respectful system)*
An action system $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G \rangle$ is called weakly respectful if and only if, for every $w \in \mathcal{D}$, every $e \in \mathcal{E}$, w is respected by every pair r_i, r_{i+1} in every transition chain **for the state** *w, and the last element of the chain is a member of $\mathcal{D}$.*

The selection function of the weakly respectful action system is given by

$$Res_{C\mathcal{D}G}(w,e) = \{r_k \in \mathcal{D} : G(w,e,r_1), C(r_i,r_{i+1}), \lhd_w(r_i,r_{i+1}), 1 \leq i < k\}.$$

The difference between this selection function and the one given in Section 6.1 is, of course, the requirement that a successor state belongs to $\mathcal{D}$ and not just to the set of stable states $\mathcal{S}_c$, and the condition that any pair (r_i, r_{i+1}) in a transition chain from the state r_1 satisfying the direct effects of the action to a successor state r_k respects the initial state w.

We also define a *trivial system* as a system where invocation never connects with causality. In other words, there are no transition chains in a trivial action system.

Definition 6.4.3 *(Trivial system)*
An action system $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G\rangle$ is called trivial, if and only if for any states r, p and an action $e \in \mathcal{E}$ such that $G(e,r,p)$, there is no state q such that $C(p,q)$.

The following two observations shed more light on the nature of respectfulness.

Lemma 6.4.4 *A strongly respectful system is trivial.*

Obviously, the use of trivial action systems is severely limited — a legitimate (admitted) state can become a successor state only if it is reached by the invocation relation directly. Let C^* be the transitive closure of C.

Lemma 6.4.5 *In a weakly respectful system, for any pair p, q such that $C^*(p,q)$, and state q is stable, and for every action e, there is no $G(e,q,p)$.*

This again restricts the invocation relation G — we need to disallow those members of G which associate the last state of a transition chain with the beginning (or any intermediate state) of the chain.

It appears that what is needed to avoid cycles and preserve an action's direct effects while propagating, is weak respectfulness. In other words, if there is a chain $p_1, ..., p_k, (q)$, where q is admitted, then invocation of any action at q cannot bring us to any of $p_1, ..., p_k$. But invocations from states other than q are allowed to do this.

However, absence of cycles is not the only restriction imposed by weak respectfulness. A weakly respectful system must satisfy another requirement—also concerning

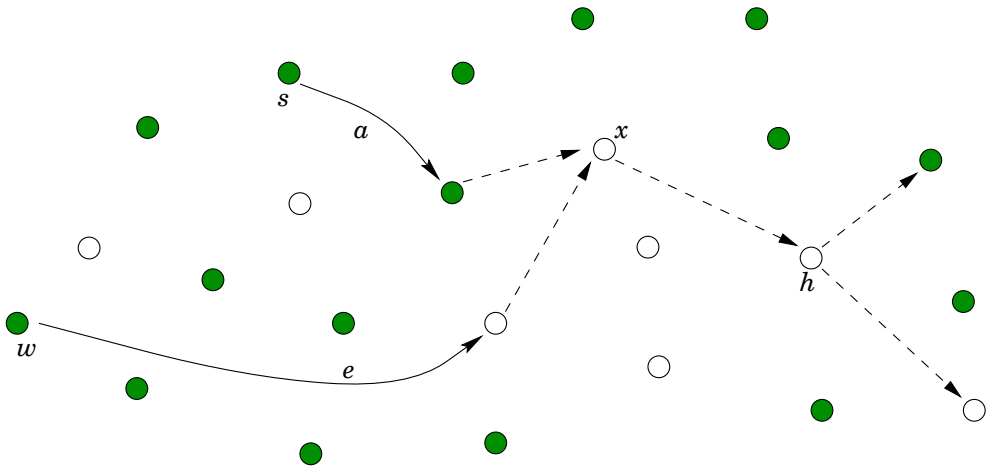

Figure 6.2: States w and s share the causal link (x, h).

the invocation relation G. If transition chains intersect then, due to the transitivity of the relation C^*, some causal links become shared by chains. Since these shared links have to respect all the initial states r, where an action invocation $G(e, r, p_1)$ propagates along more than one transition chain, an additional restriction on G must be enforced.

We say that two states w and s share a causal link (p_1, p_2) if and only if the pair (p_1, p_2) belongs to transition chains for w and s (Figure 6.2).

Lemma 6.4.6 *In a weakly respectful system, any two states w and s that share a causal link (p_1, p_2), agree on all state variables f that change values between p_1 and p_2: $p_1(f) \neq p_2(f)$ implies $w(f) = s(f)$.*

Clearly, when transition chains intersect, more chains are formed by combining beginning and ending sub-chains. And any link in ending sub-chains is shared. Therefore, the longer the shared propagation is, the fewer the number of states allowed to participate in it. The following lemma captures this intuition formally.

Lemma 6.4.7 *In a weakly respectful system where every state has n state variables, the number of states allowed to share causal chains of length k is restricted from above by $2^{(n-k)} - 1$.*

Another side of the observed dependency is that if s and w are complementary states — in other words, $w(f) \neq s(f)$ for all state variables f — then these two states cannot share any causal link.

These connections and lemma 6.4.1 indicate that, in a weakly respectful action system, the binary relation C is not a component independent from the invocation relation G. To make these components independent, in the context of a general semantics, one would need either to relax the requirement of weak respectfulness, or further constrain orderings $<_w$ in order to capture *weakly respectful* action systems. The following two conditions will prove to be useful.

(M_4) if $p \prec_w q$ then $p <_w q$.

The additional condition (M_4) ensures that an ordering $<_w$ incorporates the PMA ordering, or, in other words, includes all pairs p, q such that $p \prec_w q$. It is easy to verify that the following observation follows from the condition (M_4).

Corollary 6.4.8 *If an ordering $<_w$ satisfies (M_4) then*

$$min(<_w, [e]) \subseteq min(\prec_w, [e]).$$

Proof:

Let $r \in min(<_w, [e])$. We need to show that $r \in min(\prec_w, [e])$.

By definition of $min(<_w, [e])$, there are no states x in $[e]$ different from r such that $x <_w r$. In other words, for all states x in $[e]$ different from r, $x \not<_w r$. Then, the condition (M_4) ensures that $x \not\prec_w r$. Therefore, by definition of $min(\prec_w, [e])$, $r \in min(\prec_w, [e])$. Hence,

$$min(<_w, [e]) \subseteq min(\prec_w, [e]).$$

■

Our next condition corresponds to the Unique Minimality assumption [44], and is similar to what Lewis calls *Stalnaker's assumption* [26].

(M_5) For every action e and state w, the set $min(\prec_w, [e])$ is a singleton.

The condition (M_5) relates to a connectivity of the set $[e]$ in terms of an ordering $\prec_w$. In other words, it ensures that the minimal element of the set $[e]$ is preferred to any other element of $[e]$. Otherwise, if there are, for example, two minimal elements p and q in the set $min(\prec_w, [e])$, then they are obviously incomparable in terms of $\prec_w$. That would be a potential obstacle on the way to a weakly respectful action system simply because there might be a causal link $C(p, q)$ in some transition chain for the state w.

The next chapter will finally introduce a general augmented preferential semantics, capable of capturing *weakly respectful* action systems, while preserving sets of intended models. The conditions (M_4) and (M_5) will be shown to be sufficient to ensure that propagation along a transition chain starts from an element of $min(\prec_w, [e])$, and links states that pair-wise respect the initial state (i.e., for each causal link $C(p_i, p_j)$ in the chain, the preference $p_i \prec_w p_j$ holds). At this stage, however, the following simple result illustrates sufficient conditions for weak respectfulness. Let C' denote a *stratified* causal relation, extracted from a given relation C by preserving only links connecting to stable states: $C'(p, q)$ if and only if $C^*(p, q)$ and $q \in \mathcal{S}_c$.

Lemma 6.4.9 *If an ordering $<_w$ satisfies conditions $(M_1)-(M_5)$ for each state $w \in \mathcal{W}$, and $G_<$ is the invocation relation constructed by Definition 6.2.4 from orderings $<_w$, then an action system $\langle \mathcal{W}, \mathcal{E}, C', \mathcal{D}, G_< \rangle$, where C' is a stratified causal relation, is weakly respectful.*

Proof:

Let an ordering $<_w$ satisfy conditions $(M_1) - (M_5)$ for each state $w \in \mathcal{W}$, and $G_<$ be the invocation relation constructed by definition 6.2.4 from orderings $<_w$.

Then it is clear that the invocation relation $G_<(e, w, r)$ will choose (by its definition) precisely $<_w$-minimal states in $[e]$ for every invoked action e. We need to show that state w is respected by any pair of states in the transition chains originating from r.

From Corollary 6.4.8

$$min(<_w, [e]) \subseteq min(\prec_w, [e]).$$

Moreover, the set of minimal states $min(\prec_w, [e])$ is a singleton $\{r\}$ (by condition (M_5)). Hence, we obtain that any transition chain must start in the $\prec_w$-minimal state $r \in [e]$. It

also must end in a stable state $q \in [e]$ (given the stratified causal relation $C'(r, q)$). Since $r \in min(\prec_w, [e])$ and is the only $\prec_w$-minimal state, and $q \in [e]$ as well, it follows that $r \prec_w q$. Therefore, $\lhd_w(r, q)$ (by Lemma 6.4.1). This holds for every transition chain for the state r, ensuring that the whole action system is weakly respectful.

■

This lemma is needed to show selection-equivalence between action systems based on the causal propagation semantics and the general augmented preferential semantics.

6.5 Summary

The focus of this chapter was to demonstrate that, under certain conditions, the Principle of Minimal Change can be found in the invocation relation used by the causal propagation semantics. The presented findings support the view that minimal change and causality co-exist in separate roles and enhance each other in Sandewall's framework as well. More precisely, the causal propagation semantics with the restricted invocation relation is compatible with the augmented preferential semantics. This comparison was achieved under the approximation $\mathcal{W} = \Gamma$.

We should point out, however, that restrictions imposed on the invocation relation G in the causal propagation semantics may seem to limit the possible preconditions of actions. We believe that Sandewall has chosen a more general definition of invocation because direct effects were not the topic of his study on ramifications. Arguably, an action's direct effects should not be contingent on the state of invocation. However, in some domains such specificity may be required. In these cases, instead of restricting the invocation relation G and orderings in $\mathcal{O}$, we may choose to abandon the approximation $\mathcal{W} = \Gamma$. In other words, we may consider the information state-space Γ where relevant preconditions are properly encoded, enabling preferred propagation and achieving the desired selection-equivalence. The precise nature of a trade-off between restrictions on the preference relation and dimensionality of the propagation space remains a subject for future research.

Chapter 7

General Augmented Preferential Semantics

A unifying semantic framework for different reasoning approaches provides an ideal tool to compare these competing alternatives. A historic example is Shoham's work on *preferential* semantics [60] which provided a much needed framework to uniformly represent and compare a variety of nonmonotonic logics (including some logics of action). However, as has been shown in previous chapters, a pure preferential semantics alone is not capable of providing such a unifying framework. On the other hand, variants of preferential semantics augmented by additional structures on the state space were successfully used to characterise three influential approaches to reasoning about action and causality. The primary aim of this chapter is to provide an augmented preferential semantics that is general enough to subsume the considered frameworks to reasoning about action and causality — McCain-Turner's proposal based on causal fixed-points [37], Thielscher's causal relationships approach [63] and Sandewall's causal propagation semantics [56].

The approaches considered in previous chapters argued for an explicit representation of causal information as a way to solve the Frame and Ramification problems in a concise manner. We have shown that these approaches demand a more complex semantics than pure preferentiality can supply. For example, McCain and Turner's causal theory of action [37] was characterised by the augmented preferential semantics, using an appropriately constructed binary relation on states in addition to a preference relation (Chapter 4). Thielscher's [63] causal theory of action was characterised by another variant of an augmented preferential semantics, where the minimality component was complemented

by a binary causal relation on information states of higher dimension (Chapter 5). We have also shown in Chapter 6 that another rather general semantical approach — the causal propagation semantics proposed by Sandewall [56] — utilises the Principle of Minimal Change (hidden behind action invocation) as well, complemented by causal propagation in the state-space.

This chapter generalises the augmented preferential semantics, introduced in Chapter 2. The final variant is general enough to subsume all mentioned frameworks to reasoning about action and causality — McCain-Turner's proposal based on causal fixed-points [37], Thielscher's causal relationships approach [63], and (under certain conditions) Sandewall's causal propagation semantics [56]. The main technical contribution of this chapter is the identification of a specific semantical component — a family of choice functions — required to represent context-sensitivity of causal propagation.

7.1 Context-sensitive Propagation

A simple variant of the augmented preferential semantics was presented in Chapter 2 (section 2.6). Here, we enhance it with a component responsible for capturing context-sensitive propagation observed in the considered approaches. This new component Σ is a family of choice functions defined on $\Gamma \times \mathcal{E} \times \Gamma$, and pin-pointing *gradient* areas in 2^{Γ} — the "slopes" in the information space from which the process of causal propagation starts.

7.1.1 General Semantics: Minimal Change vs Causal Change

The *general augmented preferential semantics* can be presented as a tuple

$$\mathcal{H} = \langle \mathcal{W},\ \mathcal{D},\ \Gamma,\ \mathcal{E},\ \mathcal{O},\ \mathcal{M},\ \Sigma \rangle,$$

where

- $\mathcal{W}$ is the set of world states;
- $\mathcal{D}$ is the set of legitimate world states, $\mathcal{D} \subset \mathcal{W}$;
- Γ is the set of information states;

- $\mathcal{E}$ is the set of actions;
- $\mathcal{O}$ is the preferential structure on $\Gamma \times \Gamma$ (the set of orderings $<_\alpha$ defined with respect to each information state $\alpha \in \Gamma$);
- $\mathcal{M}$ is the causal binary relation on $\Gamma \times \Gamma$;
- Σ is the family of choice functions $\sigma : \Gamma \times \mathcal{E} \times \Gamma \to 2^\Gamma$.

together with the functions

- post-condition $[e] : \mathcal{E} \to 2^\mathcal{W}$;
- projection $\mathcal{P}(\gamma) : \Gamma \to \mathcal{W}$;
- selection $Res(w, e) : \mathcal{W} \times \mathcal{E} \to 2^\mathcal{W}$.

In other words, the Principle of Minimal Change (identified with $\mathcal{O}$) and the Principle of Causal Change (identified with $\mathcal{M}$) are explicitly represented and play a clear and distinct role. We maintain that both principles are required to solve the Frame and Ramification problems in a concise fashion.

To elaborate on this, let us re-iterate the technical process of obtaining a successor state, given an initial state and an action — motivated by a simple and clear state transition semantics. Intuitively, a successor state is an admitted state which is reachable (by means of some transition relation) from states nearest to the initial one among states satisfying the post-conditions of the performed action. In other words, tracing all effects of the action (both direct and indirect) may involve two steps. First, we find states satisfying the action's post-conditions while staying as close as possible to the initial state according to some preference relation. Then, we propagate along the transition relation from all such minimal states all the way to some stable (and admitted) state — our result state.

However, sometimes there is a causal context present in the domain that may require additional checks and balances, potentially obscuring this simple and clear view. For example, the causal relationships approach introduced by Thielscher [63] attempts to encode contextual information in (state, history) pairs while tracing causal ramifications.

To capture context-sensitivity, a bounded start area (a gradient) is chosen in the information state-space Γ. This can be done using an ordering $<_\gamma \in \mathcal{O}$ and a choice function $\sigma \in \Sigma$. As mentioned earlier (Chapter 2, Section 2.6), the agent entertains the states in the gradient area as the most preferred information states compatible with the action's direct effects. Intuitively, the purpose of the information states is to serve as possible causal extensions of normal world states, providing necessary context to the process of causal propagation. The propagation starts from the gradient and is driven (in a very simple way) by the causal relation $\mathcal{M}$. This propagation may explore the whole state-space Γ, but is expected to reach at least one stable information state.

The difference between the simple variant sketched in Chapter 2 and the general augmented preferential semantics is that a final stable information state must be $\mathcal{M}$-reachable from all *the states in the gradient area.*

As with the simple variant of our semantics, if a projection from such a final state to the state-space $\mathcal{W}$ results in a legitimate state r compatible with e (in other words, $r \in \mathcal{D} \cap [e]$), then the state r is the desired successor state of the action at hand.

At this stage, we need to repeat some notation from Chapter 2, and in particular, from section 2.3 — summarised at the end of Chapter 2.

The set $[e]^\Gamma$ is the set of information states whose normal-space projections make up the post-condition set $[e]$. Similarly, the set $\mathcal{D}^\Gamma$ contains all information states that would project to the admitted states in $\mathcal{D}$. Accordingly, the set $min(<_\gamma, [e]^\Gamma)$ is a subset of $[e]^\Gamma$ containing states nearest to the state γ in terms of the ordering $<_\gamma$. In other words,

$$min(<_\gamma, [e]^\Gamma) = \{\beta \in [e]^\Gamma, \neg\exists\alpha \in [e]^\Gamma, \alpha \neq \beta, \alpha <_\gamma \beta\}.$$

An element of $min(<_\gamma, [e]^\Gamma)$ can be referred to as a state $<_\gamma$-minimal in $[e]^\Gamma$.

The set-projection function $\mathcal{X}$ maps sets of information states onto sets of world states from $\mathcal{W}$. In other words, the function $\mathcal{X} : 2^\Gamma \rightarrow 2^\mathcal{W}$ is defined as follows: $\mathcal{X}(\{\gamma_1, ..., \gamma_n\}) = \{\mathcal{P}(\gamma_1)\} \cup ... \cup \{\mathcal{P}(\gamma_n)\}$, where the function $\mathcal{P}$ is our projection function.

The important density condition is given by

$$\mathcal{D} \cap \mathcal{X}(\Gamma \setminus \mathcal{K}_\mathcal{M}) = \emptyset,$$

where $\mathcal{K}_{\mathcal{M}}$ is the set of stable information states. This condition basically specifies that an unstable information state should not be projected onto a legitimate state. It implies that

$$\mathcal{D} \subseteq \mathcal{X}(\mathcal{K}_{\mathcal{M}}).$$

In other words, as noted in Chapter 2, some domain constraints may eliminate more illegitimate states than causal propagation alone. Arguably, any state which is not admitted, should be excluded by some causal laws—and the set $\mathcal{D}$ can be *made* equal to the set $\mathcal{X}(\mathcal{K}_{\mathcal{M}})$ of stable states. However, the distinction between admitted and (causally-)stable states may be useful in providing some flexibility to domain constraints. For example, stable states may reflect hard constraints, while admitted states may correspond to soft constraints. Varying selection requirements on successor states (stable and admitted, or merely stable) would allow us to capture an additional degree of state eligibility if required.

Before specifying a selection function Res for the general augmented preferential semantics, we need to (re-)define a few more (familiar) constructs derived from elements of the action system $\mathcal{H}$.

Firstly, we define *e-predecessors* of a given information state — a set of information states preceding a given state with respect to an ordering from $\mathcal{O}$. Formally:

Definition 7.1.1 *(e-predecessors)*
Given any two information states γ, β and any action e, the set of e-predecessors of β with respect to γ is defined to be the set

$$\langle\!|\beta, e|\!\rangle_{\gamma} = \{\alpha : \ \alpha \in [e]^{\Gamma} \textit{ and } \alpha <_{\gamma} \beta\}.$$

The e-predecessors of β with respect to γ are just the $[e]^{\Gamma}$ states which are closer to γ than β. The ordering $<_{\gamma}$ is reflexive, and it is clear that any $\beta \in [e]^{\Gamma}$ is an e-predecessor of itself with respect to γ (in other words, $\beta \in \langle\!|\beta, e|\!\rangle_{\gamma}$).

It is interesting at this stage to compare the states $<_{\gamma}$-minimal in $[e]^{\Gamma}$, and the e-predecessors of some information state β with respect to γ (Figure 7.1). Intuitively, the $<_{\gamma}$-minimal states in $[e]^{\Gamma}$ compose a boundary separating the set $[e]^{\Gamma}$ from an information state corresponding to the initial normal state w. A projection $\mathcal{P}$ of a $<_{\gamma}$-minimal

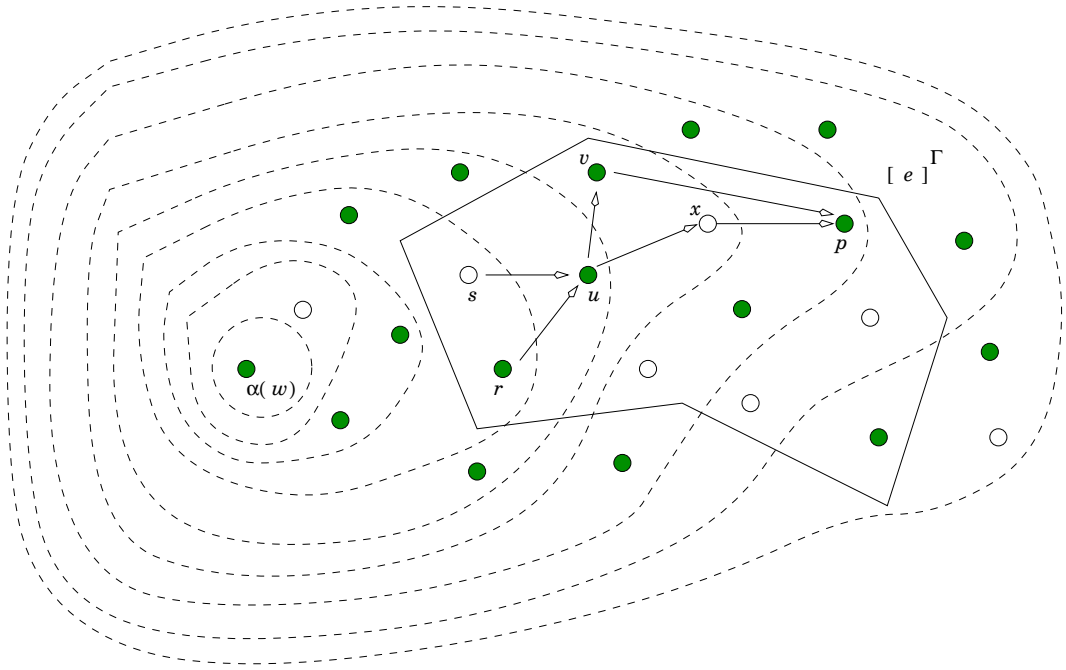

Figure 7.1: The e-predecessors of the information state p include states s, r, u, v, x and p. The direct preferences are shown as arrows. The $<_{\alpha(w)}$-minimal states in $[e]^\Gamma$ include states s and r.

information state results in an intermediate normal state that is used to form a starting point of propagation in Thielscher's and Sandewall's approaches. In other words, the boundary $min(<_\gamma, [e]^\Gamma)$ is the earliest "zone" where we may start causal propagation in the information space. In some sense, the states on the boundary "support" the propagation, and can be thought of as "collaborators" in producing successor state(s). On the other hand, the e-predecessors in $\langle\!\langle \beta, e \rangle\!\rangle_\gamma$ "compete" against each other because any of them could be eventually selected as a successor information state. For example, in the McCain-Turner approach, all predecessors of a causal fixed-point r have to be eliminated during the state transition process leading to r — so, in a sense, the e-predecessors in $\langle\!\langle \beta, e \rangle\!\rangle_\gamma$ frame a potentially final point (or a *horizon*) of propagation.

By definition, some e-predecessors of an information state β with respect to γ, lie on the boundary $min(<_\gamma, [e]^\Gamma)$:

$$min(<_\gamma, e) \cap \langle\!\langle \beta, e \rangle\!\rangle_\gamma \neq \emptyset.$$

Intuitively, the sets $min(<_\gamma, e)$ and $\langle\!\langle \beta, e \rangle\!\rangle_\gamma$ represent a minimality-driven component of an action system, and indicate a possible *gradient* of causality-driven propagation (a "slope" in the information space along which the propagation begins). In other words, the minimal change component contributes to the selection function by shaping the state-space "boundaries" and surfaces for *coordinated* causal propagation.

7.1.2 General Semantics: Gradient Choice Functions

The motivation behind our general augmented preferential semantics is to represent various approaches uniformly. That is why an attempt is made to accommodate seemingly different selection functions discussed in chapters 4, 5 and 6 in one generic selection function Res. To achieve this goal, however, we will need to employ different choice functions from Σ aimed at identifying (narrowing down) the minimality-driven "gradient" of state transitions. A choice function is defined for a potential successor information state $\beta \in \Gamma$, an action $e \in \mathcal{E}$, and an information state γ such that its projection is a given initial state $w \in \mathcal{W}$ (in other words, $\mathcal{P}(\gamma) = w$). We shall call states chosen by a choice function $\sigma(\beta, e, \gamma)$, the σ-chosen states in the (β, e, γ)-gradient.

Our first choice function $\sigma_F(\beta, e, \gamma)$, called a *full-meet gradient*, merely returns all e-predecessors of some information state β with respect to γ:

$$\sigma_F(\beta, e, \gamma) = \langle\!\langle \beta, e \rangle\!\rangle_\gamma$$

The second choice function $\sigma_M(\beta, e, \gamma)$, called a *mini-choice gradient*, chooses one $<_\gamma$-minimal state in $[e]^\Gamma$, or in other words, an element of $min(<_\gamma, [e]^\Gamma)$:

$$\sigma_M(\beta, e, \gamma) = \{\alpha\}, \text{ where } \alpha \in min(<_\gamma, [e]^\Gamma)$$

The full-meet gradient defines a set $\sigma_F(\beta, e, \gamma)$ that identifies potential challengers for a successor information state place. Intuitively, the states in this set compete to become a successor information state and the information state β is being tested as a potential winner.

The mini-choice gradient, on the contrary, does not test any potential winner or challengers (in fact, it is independent of the first argument). Instead, this choice function attempts to make subsequent causal propagation as flexible as possible — by picking just

one of the $<_\gamma$-minimal states in $[e]^\Gamma$. Intuitively, the "mini-choice gradient" supports all possibilities in terms of intermediate information states consistent with the direct effects of an action (more precisely, projections of these intermediate states satisfy the direct effects of an action).

Sandewall's causal propagation semantics, for example, allows the propagation to start at any such intermediate state — accounting, in particular, for disjunctive direct effects. Thielscher's approach also suggests to start state transitions at an intermediate state that is as close as possible to the initial state. Since the latter approach handles actions with conjunctive effects, there is only one such intermediate state. However, an approach reported by Thielscher [64] extends the original one [63] towards *alternative* effect propositions, where the disjunction of effects $e_1 \vee e_2$ is interpreted as exclusive, and inclusive disjunction is simply modelled as $e_1 \vee e_2 \vee (e_1 \wedge e_2)$. In this extended case, alternative effects lead to alternative intermediate (preliminary) states, and the original causal relationship approach is then applied to each of these preliminary states. In other words, in order to account for indirect effects, preliminary states are "taken as starting points for the successive applications of causal relationships" until overall satisfactory successor states are obtained [64].

Therefore, the mini-choice gradient should be considered if an action system includes non-deterministic actions, or more precisely, actions with disjunctive direct effects.

We introduce two more choice functions. One, $\sigma_P(\beta, e, \gamma)$, is called a *partial-choice gradient* [1], and chooses a $<_\gamma$-minimal state in $[e]^\Gamma \cap \mathcal{D}^\Gamma$ or in other words, a minimal element among information states that would project onto admitted normal states consistent with the direct effects of action e.

$$\sigma_P(\beta, e, \gamma) = \{\alpha\}, \text{ where } \alpha \in min(<_\gamma, [e]^\Gamma \cap \mathcal{D}^\Gamma)$$

It is important to point out that the partial-choice gradient $\sigma_P(\beta, e, \gamma)$ is weaker than a more demanding choice function

$$\sigma_L(\beta, e, \gamma) = \{\alpha\}, \text{ where } \alpha \in min(<_\gamma, [e]^\Gamma) \cap \mathcal{D}^\Gamma$$

[1] It can be easily guessed that names of some choice functions that we use are loosely based on well-known *full-meet*, *partial meet* and *maxi-choice* belief revision functions [13].

which could be referred to as a *legitimate-choice gradient.*

Intuitively, the legitimate-choice gradient chooses a $<_\gamma$-minimal state in $[e]^\Gamma$ that is, in addition, legitimate with respect to $\mathcal{D}$ (more precisely, its projection is in $\mathcal{D}$). It may result, sometimes, in the empty set — when all information states on the boundary $min(<_\gamma, [e]^\Gamma)$ are not legitimate with respect to $\mathcal{D}$.

On the contrary, the partial-choice gradient function focuses directly on information states in $min(<_\gamma, [e]^\Gamma \cap \mathcal{D}^\Gamma)$. Clearly,

$$min(<_\gamma, [e]^\Gamma) \cap \mathcal{D}^\Gamma \subseteq min(<_\gamma, [e]^\Gamma \cap \mathcal{D}^\Gamma)$$

The set on the right-hand side of the containment can be thought of as another information space boundary, "located" further from the initial point γ than the boundary $min(<_\gamma, [e]^\Gamma)$, and the "partial-choice gradient" function targets these information states.

It is worth pointing out again that all the gradient choice functions narrow down the search and selection of potential (context-sensitive) successor states in the information space, based on a preference relation $<_\gamma$, and without the causality component $\mathcal{M}$ of an action system. In other words, the gradient choice functions scale potential causal propagation towards context-sensitive ramifications.

7.1.3 General Semantics: Selection Function

Let $\mathcal{M}^*$ be the transitive closure of the relation $\mathcal{M}$. As usual, we say that an information state β is $\mathcal{M}$-reachable from an information state α, if $\mathcal{M}^*(\alpha, \beta)$. Also, for simplicity, we denote by γ_w any information state such that $\mathcal{P}(\gamma_w) = w$. Now we are ready to define our final selection function.

We say that an admitted state r is a successor state, $r \in Res(w, e)$, if and only if r is a projection of some stable information state β, which is $\mathcal{M}$-reachable from *all* σ-chosen information states in the (β, e, γ_w)-gradient. More precisely,

Definition 7.1.2 *(Selection Function) A selection function of an action system* $\mathcal{H} = \langle \mathcal{W}, \mathcal{D}, \Gamma, \mathcal{E}, \mathcal{O}, \mathcal{M}, \Sigma \rangle$ *is given as*

$$Res(w,e) = \{r \in \mathcal{D} : \textit{ for all } \alpha \in \sigma(\beta, e, \gamma_w),\ \mathcal{M}^*(\alpha,\beta), \\ \textit{where } \beta \in \mathcal{K}_\mathcal{M} \textit{ and } \mathcal{P}(\beta) = r\}.$$

The notions of "stable" and "reachable" are understood in terms of the causal transition relation $\mathcal{M}$, and the gradient is given by one of the choice functions, using an ordering $<_{\gamma_w}$ from $\mathcal{O}$.

This definition does not require that successor states $Res(w,e)$ satisfy the direct effects of an action. In other words, we did not require $Res(w,e) \subseteq [e]$. However, this property may be seen as one of the most fundamental to selection of successor states. This motivates us to introduce a sub-class of *conservative* action systems, where any successor state must satisfy the direct effects of actions.

Definition 7.1.3 *A selection function of a* $conservative$ *action system* $\mathcal{H} = \langle \mathcal{W},\ \mathcal{D},\ \Gamma,\ \mathcal{E},\ \mathcal{O},\ \mathcal{M},\ \Sigma\rangle$ *is given as*

$$Res(w,e) = \{r \in \mathcal{D} \cap [e] : \textit{ for all } \alpha \in \sigma(\beta, e, \gamma_w),\ \mathcal{M}^*(\alpha,\beta), \\ \textit{where } \beta \in \mathcal{K}_\mathcal{M} \textit{ and } \mathcal{P}(\beta) = r\}.$$

We shall see that action systems capturing McCain-Turner and Sandewall frameworks are in fact conservative.

If we believe that $\mathcal{D} = \mathcal{X}(\mathcal{K}_\mathcal{M})$, or in other words, a distinction between stable and admitted states is not required, then the selection function can be accordingly simplified. This is the sub-class of *compact* action systems where domain constraints are captured by causal dependencies alone, embodied in the causal transition relation $\mathcal{M}$:

Definition 7.1.4 *A selection function of a* $compact$ *action system* $\mathcal{H} = \langle \mathcal{W},\ \mathcal{D},\ \Gamma,\ \mathcal{E},\ \mathcal{O},\ \mathcal{M},\ \Sigma\rangle$ *is given as*

$$Res(w,e) = \{r \in \mathcal{W} : \textit{ for all } \alpha \in \sigma(\beta, e, \gamma_w),\ \mathcal{M}^*(\alpha,\beta), \\ \textit{where } \beta \in \mathcal{K}_\mathcal{M} \textit{ and } \mathcal{P}(\beta) = r\}.$$

Obviously, there could be *conservative* and *compact* action systems, where the selection function is given as

$$Res(w,e) = \{r \in [e] : \text{ for all } \alpha \in \sigma(\beta, e, \gamma_w),\ \mathcal{M}^*(\alpha, \beta), \\ \text{where } \beta \in \mathcal{K}_\mathcal{M} \text{ and } \mathcal{P}(\beta) = r\}.$$

We shall specify another important sub-class of action systems, where the selection function can be modified; more precisely, strengthened as follows.

Definition 7.1.5 *A selection function of a $Hamiltonian$ action system $\mathcal{H} = \langle \mathcal{W},\ \mathcal{D},\ \Gamma,\ \mathcal{E},\ \mathcal{O},\ \mathcal{M},\ \Sigma \rangle$ is given as*

$$Res(w,e) = \{r \in \mathcal{D} : \textit{ for all } \alpha \in \sigma(\beta, e, \gamma_w),\ \mathcal{M}^*(\alpha, \beta), \\ \textit{there is a Hamiltonian path through states in } \sigma(\beta, e, \gamma_w), \\ \textit{where } \beta \in \mathcal{K}_\mathcal{M} \textit{ and } \mathcal{P}(\beta) = r\}.$$

In other words, not only should a stable successor information state be reachable from *all* information states lying on the chosen gradient, but there must be a Hamiltonian path through these states. When a successor information state is itself on the gradient (for example, "full-meet gradient"), the requirement of a Hamiltonian path is clearly stronger than the requirement of reachability. This is due to the condition that a successor information state should be stable, which ensures that such a Hamiltonian path ends in this information state. Therefore, this state is reachable from all gradient states.

A Hamiltonian conservative compact action system will be required to show selection-equivalence with McCain-Turner action systems.

7.2 Examples of Reduction

7.2.1 Preferential Semantics

Before we demonstrate how the desired selection-equivalence can be achieved for all three approaches considered previously, we consider a most obvious simplification of action systems, captured by the general augmented preferential semantics $\mathcal{H}$. Staying within normal state-space ($\mathcal{W} = \Gamma$, $\mathcal{P}(\iota) = \iota$), setting $\mathcal{M} = \emptyset$ and taking the "partial-choice gradient" $\sigma_P(\beta, e, \gamma)$ as our choice function, produces a traditional preferential semantics with a variety of suitable preference relations $\mathcal{O}$. More precisely, the selection function is given then as

$$Res_P(w,e) = \{r \in min(<_w, [e] \cap \mathcal{D})\}.$$

Obviously, taking a more demanding "legitimate-choice gradient" $\sigma_L(\beta, e, \gamma)$ as our choice function, produces a preferential semantics tending to disqualify more successor states:

$$Res_L(w,e) = \{r \in \mathcal{D} \cap min(<_w, [e])\}.$$

Clearly, action systems based on these selection functions are conservative.

7.2.2 Causal Systems with Fixed-points

Now we shall specify action systems within the general augmented preferential semantics $\mathcal{H}$ that capture our motivating approaches.

We begin with our semantics for the McCain-Turner framework. As mentioned above, we focus on Hamiltonian conservative compact action systems. In this case, we do not need the information space concept, as causal context will be represented entirely by a Hamiltonian path through e-predecessors of a successor state. Therefore, we shall stay within normal state-space ($\mathcal{W} = \Gamma$, $\mathcal{P}(\iota) = \iota$) and set all orderings in $\mathcal{O}$ to be PMA orderings, as defined in Chapter 4. The transition relation $\mathcal{M}$ is constructed as described in Chapter 4 as well. Since we do not extend to information space, some notation simplifies as well. For example, by definition, $[e]^\Gamma = [e]$, and $\mathcal{K}_\mathcal{M} = \mathcal{X}(\mathcal{K}_\mathcal{M})$.

Importantly, we take the "full-meet gradient" $\sigma_F(\beta, e, \gamma)$ as our choice function. This results in the following reduction:

$$\begin{gathered} Res(w,e) = \{r \in \mathcal{D} \cap [e] : \text{ for all } \alpha \in \langle\!\langle r, e]\!]_w, \ \mathcal{M}^*(\alpha, r), \\ \text{there is a Hamiltonian path through states in } \langle\!\langle r, e]\!]_w, \\ \text{and } r \in \mathcal{K}_\mathcal{M}\}, \end{gathered}$$

Given the compactness of the action system, $\mathcal{D} = \mathcal{X}(\mathcal{K}_\mathcal{M}) = \mathcal{K}_\mathcal{M}$, and the fact that a successor state is in the chosen gradient (and therefore, is reachable from all states in the gradient), we can simplify the selection function further:

$$\begin{gathered} Res(w,e) = \{r \in \mathcal{K}_\mathcal{M} \cap [e] : \\ \text{there is a Hamiltonian path through states in } \langle\!\langle r, e]\!]_w\}. \end{gathered}$$

Re-iterating, a stable state r satisfying the direct effects of the action e is a successor state if and only if there is a Hamiltonian path through all its e-predecessors (this path is bound to end in r).

This is precisely the selection function $\mathtt{Succ}_{\mathcal{M}}(w, e)$ used in Chapter 4 to achieve the desired selection-equivalence with the original McCain-Turner characterisation of causal fixed-points.

Therefore, the following result (based on Corollary 4.6.6 of Chapter 4) can be obtained.

Theorem 7.2.1 *For every causal action system $\mathcal{Q}$ there exists a selection- equivalent Hamiltonian conservative compact action system $\langle \mathcal{W},\ \mathcal{D},\ \Gamma,\ \mathcal{E},\ \mathcal{O},\ \mathcal{M},\ \Sigma\rangle$.*

Conversely, for every Hamiltonian conservative compact action system $\langle \mathcal{W},\ \mathcal{D},\ \Gamma,\ \mathcal{E},\ \mathcal{O},\ \mathcal{M},\ \Sigma\rangle$, with $\mathcal{W} = \Gamma$ and the full-meet choice function, there exists a selection-equivalent causal action system $\mathcal{Q}$.

7.2.3 Systems with Causal Relationships

The next example is, in fact, the only one that uses the information space concept. Nevertheless, we believe that it exemplifies quite a general class of action systems. We refer here to the framework of Thielscher analysed in Chapter 5.

In order to achieve the desired selection-equivalence, we reproduce the information space Γ as described in Chapter 5. We also take $\mathcal{M} = \rightharpoonup$, and include in the set $\mathcal{E}$ all the actions a from the set of action laws A, where each action law has the form $\langle C, a, E\rangle$, E being the direct effect of a.

The set $\mathcal{D}$ is chosen as a subset of $\mathcal{W}$ such that its elements satisfy the constraints D.

The projection function $\mathcal{P}$ is defined via the hyper-space projection function p as follows:

$$\mathcal{P}(\gamma(z)) = p(s),$$

where $\gamma(z)$ is an information (power-) state, z is the corresponding set of hyper-states, and $s \in z$ is an arbitrary element.

For example, consider the state $q = \{a, \neg b, c\}$ and two hyper-states $s = \{a, \neg b, c, \overset{\circ}{a}, \neg\overset{\circ}{b}, \overset{\circ}{c}\}$ and $s' = \{a, \neg b, c, \neg\overset{\circ}{a}, \neg\overset{\circ}{b}, \neg\overset{\circ}{c}\}$. Let $z = \{s, s'\}$. Then the information (power-)

state $\gamma(z)$ corresponds to the partial state $\gamma_q(z) = \{a, \neg b, c, \neg\overset{\circ}{b}\}$. The projection of $\gamma(z)$ to the space $\mathcal{W}$ simply returns the state q.

The next step is a construction of the preferential structure $\mathcal{O}$. The intention is to construct orderings $\ll$ in $\mathcal{O}$ in such a way that $\ll$-minimal elements would be pinpointed by appropriate choice functions. Importantly, the construction of any $\ll_\gamma$ should avoid any references to actions in $\mathcal{E}$ — in other words, the preferential structure $\mathcal{O}$ should be action-independent.

We intend to base the preferential structure in the information space Γ on the PMA ordering, at least in part. Informally, given three information states γ_w, γ_1 and γ_2 that correspond to partial hyper-states z, z_1 and z_2 from different hyper-neighbourhoods $N(w)$, $N(q_1)$ and $N(q_2)$, we would consider that γ_1 is closer to γ_w than γ_2, denoted $\gamma_1 \ll_{\gamma_w} \gamma_2$, if q_1 is closer to w than q_2 in terms of the PMA ordering: $q_1 \prec_w q_2$. Intuitively, if there is a preference relation between the projections of γ_1 and γ_2 we take it as an indication of some preference between the information states themselves. Similarly, if the corresponding projections are not comparable: neither $q_1 \prec_w q_2$ nor $q_2 \prec_w q_1$, we do not wish to specify any preference between γ_1 and γ_2 as well.

There is, however, a case when some additional care must be taken — the case when two information states are projected onto the same state in $\mathcal{W}$, more precisely, $\mathcal{P}(\gamma(z_1)) = \mathcal{P}(\gamma(z_2)) = q$. This may occur when both sets z_1 and z_2 belong to the same hyper-neighbourhood $N(q)$. In this case, the only distinguishing features are these sets z_1 and z_2 and the corresponding partial hyper-states $\gamma_q(z_1)$ and $\gamma_q(z_2)$.

In order to exploit these distinguishing features in defining which information state, $\gamma(z_1)$ or $\gamma(z_2)$, is closer to γ_w, where $\mathcal{P}(\gamma_w) = w$, we need to introduce a few auxiliary functions, and an auxiliary preference relation $\sqsubset$ among partial hyper-states.

Firstly, we define the *observed change*, $Obs(q, w)$, between two normal states w and q, as the set of literals in $(q \setminus w)$ expressed in terms of justifier literals. More precisely,

$$Obs(q, w) = \overset{\circ}{U}, \quad \text{where} \quad U = q \setminus w.$$

Put simply, the observed change is the literals f from the state q that are not present in the state w, expressed in terms of justifier literals $\overset{\circ}{f}$. For example, given two states $q = \{a, \neg b, c\}$ and $w = \{a, b, c\}$, the observed change is $Obs(q, w) = \{\neg\overset{\circ}{b}\}$.

Secondly, we define the *justified change*, $Just(\gamma_q(z))$, for a partial hyper-state $\gamma_q(z)$, as the set of all justifier literals in $\gamma_q(z)$. In other words, the justified change contains all justifier literals common to all hyper-states in z. More precisely,

$$Just(\gamma_q(z)) = \gamma_q(z) \setminus q.$$

For example, if $\gamma_q(z_1) = \{a, \neg b, c, \neg\mathring{b}\}$, then $Just(\gamma_q(z_1)) = \{\neg\mathring{b}\}$, and if $\gamma_q(z_2) = \{a, \neg b, c, \neg\mathring{a}, \neg\mathring{b}\}$, then $Just(\gamma_q(z_2)) = \{\neg\mathring{a}, \neg\mathring{b}\}$.

Finally, we compare these change sets using the symmetric difference ***Diff***$(x, y) = (y \setminus x) \cup (x \setminus y)$. More precisely, we define the *divergent change*, $Div(\gamma_q(z), w)$, for a partial hyper-state $\gamma_q(z)$ and state w, as the set of justifier literals that appear either in the observed change set $Obs(q, w)$ or in the justified change set $Just(\gamma_q(z))$, but not in both. More precisely,

$$Div(\gamma_q(z), w) = \textbf{\textit{Diff}}(\ Obs(q, w),\ Just(\gamma_q(z))\ ).$$

For example, if $w = \{a, b, c\}$ and $\gamma_q(z_1) = \{a, \neg b, c, \neg\mathring{b}\}$ we obtain $Div(\gamma_q(z_1), w) = \emptyset$, and for $\gamma_q(z_2) = \{a, \neg b, c, \neg\mathring{a}, \neg\mathring{b}\}$ we obtain $Div(\gamma_q(z_2), w) = \{\neg\mathring{a}\}$.

Intuitively, a partial hyper-state $\gamma_q(z_1)$ is preferred to a partial hyper-state $\gamma_q(z_2)$ (with respect to state w) if and only if the divergent change for $\gamma_q(z_1)$ is contained in the divergent change for $\gamma_q(z_2)$ (with respect to state w). Formally,

$$\gamma_q(z_1) \sqsubset_w \gamma_q(z_2) \quad \text{if and only if}$$

$$Div(\gamma_q(z_1), w) \subseteq Div(\gamma_q(z_2), w).$$

Essentially, the $\sqsubset_w$ ordering is based on the PMA idea, generalising it to sets of justifier literals. Now we are ready to construct the preferential structure on the information space. The set $\mathcal{O}$ is a set of orderings $\ll$ defined on information states in such a way that respective projections satisfy the PMA ordering, while preferring subsets with smaller divergent change within each hyper-neighbourhood. More precisely,

$$\gamma(z_1) \ll_\gamma \gamma(z_2) \quad \text{if and only if}$$
$$\text{either } \mathcal{P}(\gamma(z_1)) \prec_{\mathcal{P}(\gamma)} \mathcal{P}(\gamma(z_2)) \quad \text{or}$$
$$\text{both } \mathcal{P}(\gamma(z_1)) = \mathcal{P}(\gamma(z_2)) = q \quad \text{and} \quad \gamma_q(z_1) \sqsubset_{\mathcal{P}(\gamma)} \gamma_q(z_2).$$

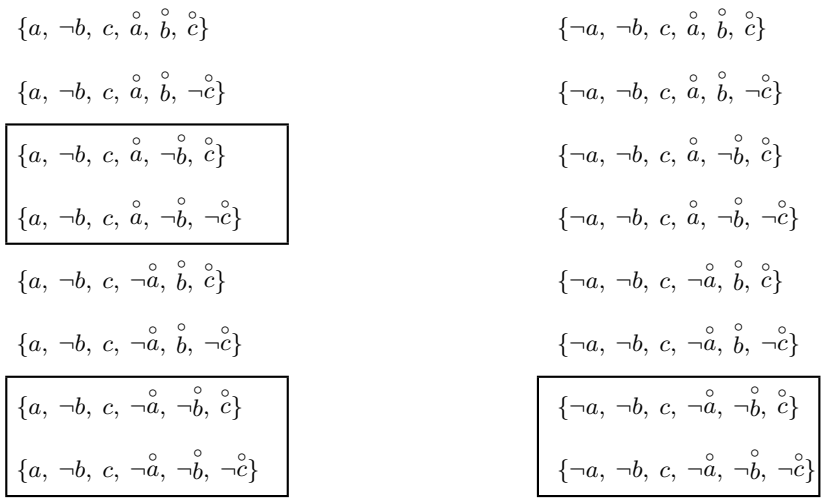

Figure 7.2: Hyper-neighbourhoods of the states $r = \{a, \neg b, c\}$ and $r' = \{\neg a, \neg b, c\}$.

In other words, an information (power-) state $\gamma(z_1)$ is nearer to γ than $\gamma(z_2)$ if and only if

- the states $\gamma(z_1)$ and $\gamma(z_2)$ correspond to different hyper-neighbourhoods, and the projection of $\gamma(z_1)$ to normal space is nearer to the projection of γ than the projection of $\gamma(z_2)$, or (if the projections are the same, i.e. the state q)
- the partial hyper-state $\gamma_q(z_1)$ is preferred to the partial hyper-state $\gamma_q(z_2)$ with respect to the projection of γ — i.e. the states $\gamma(z_1)$ and $\gamma(z_2)$ correspond to the same hyper-neighbourhood, but in the first partial hyper-state the observed change diverges from the justified change at most as much as in the second partial hyper-state.

This two-tiered PMA-based preference relation orders information (power-) states representing partial states from different hyper-neighbourhoods, and then prefers subsets with smaller divergent change within each hyper-neighbourhood.

Let us illustrate this with an example (Figure 7.2), where the set $z_1 = N(r) \cap [\neg\overset{\circ}{b}]$ includes all four hyper-states enclosed in two boxes on the left-hand side, while the set $z_2 = N(r) \cap [\neg\overset{\circ}{a} \wedge \neg\overset{\circ}{b}]$ includes two hyper-states enclosed in the bottom box on the left-hand side. The set $N(r') \cap [\neg\overset{\circ}{a} \wedge \neg\overset{\circ}{b}])$ is enclosed in the box on the right-hand side.

We again denote by w the state $\{a, b, c\}$. The state $\gamma = \gamma(N(w))$ is the information

state that corresponds to the whole hyper-neighbourhood $N(w)$. It is easy to check that $\mathcal{P}(\gamma(N(w))) = w$ (by definition of projection function), and the partial hyper-state $\gamma_w(N(w)) = w$ (by definition of partial hyper-state). We also abbreviate

$$\gamma_1 = \gamma(N(r) \cap [\neg\overset{\circ}{b}]),$$

$$\gamma_2 = \gamma(N(r) \cap [\neg\overset{\circ}{a} \wedge \neg\overset{\circ}{b}]),$$

$$\gamma_3 = \gamma(N(r') \cap [\neg\overset{\circ}{a} \wedge \neg\overset{\circ}{b}]),$$

and obtain

$$\gamma_1 \ \ll_\gamma \ \gamma_2 \ \ll_\gamma \ \gamma_3.$$

Both information states γ_1 and γ_2 that correspond to the hyper-states on the left-hand side (in sets z_1 and z_2 respectively) are preferred to the information state γ_3 corresponding to the set z_3 of hyper-states on the right-hand side. The reason is that the state r is closer to the state w than the state r' according to the PMA ordering: $r \prec_w r'$.

In addition, $\gamma_1 \ \ll_\gamma \ \gamma_2$, because the partial hyper-state $\gamma_r(z_1)$ is preferred to the partial hyper-state $\gamma_r(z_2)$ (with respect to state w):

$$\gamma_r(z_1) \sqsubset_w \gamma_r(z_2).$$

To verify the latter preference we note that the partial hyper-state $\gamma_r(z_1)$ has no divergent change from w: $Div(\gamma_r(z_1), w) = \emptyset$ (both observed and justified change sets contain only the literal $\neg\overset{\circ}{b}$), while the divergent change between w and the partial hyper-state $\gamma_r(z_2)$ is non-empty and contains the literal $\neg\overset{\circ}{a}$ (this literal belongs to the justified change set but not to the observed change set).

Not surprisingly, given an initial state w and an action with the post-condition E, the information (power-) state $\gamma(\|E\|_w)$ is a $\ll_{\gamma(N(w))}$-minimal state among all information (power-) states in $[E]^\Gamma$. We choose the partial hyper-state $\gamma_w(N(w))$ as the point of reference because it has no divergent change from w: the observed change set is empty, $(w \setminus w)$; and the justified change set is empty, $\gamma_w(N(w)) = w$. This observation is important because it identifies the information state corresponding to the trigger set as a minimal element. Therefore, this information state may become an appropriate gradient for subsequent causal propagation. Formally, we can observe the following.

Lemma 7.2.2 *For a state $w \in \mathcal{W}$ and an action law $\langle C, a, E\rangle$,*

$$\gamma(\|E\|_w) \in min(\ll_{\gamma(N(w))}, [E]^{\Gamma}).$$

We have observed earlier that Thielscher's approach handles actions with conjunctive effects, producing only one intermediate state that is as close as possible to the initial state. Hence, there is always only one corresponding trigger set, and therefore,

$$\{\gamma(\|E\|_w)\} = min(\ll_{\gamma(N(w))}, [E]^{\Gamma}).$$

Moreover, we may apply the definition of trigger sets to the extended Thielscher's approach [64] that treats *alternative* effect $E = E_1 \vee \ldots \vee E_n$. In this extended case, alternative effects lead to alternative intermediate states, and hence, to the alternative trigger sets. It is easy to show that the information states corresponding to these multiple trigger sets are $\ll_{\gamma(N(w))}$-minimal elements as well, and together make up the set $min(\ll_{\gamma(N(w))}, [E]^{\Gamma})$:

$$\{\gamma(\|E_1\|_w), \ldots, \gamma(\|E_n\|_w)\} = min(\ll_{\gamma(N(w))}, [E]^{\Gamma}).$$

In summary, we constructed the information space Γ, the preferential structure $\mathcal{O}$, the transition relation $\mathcal{M} = \rightharpoonup$, the space of legitimate states $\mathcal{D}$, the set of actions $\mathcal{E} = A$, and the projection function $\mathcal{P}$.

In addition to these constructs, we take the mini-choice gradient $\sigma_M(\beta, a, \gamma)$ as our choice function — one that chooses a $\ll_\gamma$-minimal state in $[E]^{\Gamma}$, or in other words, an element of $min(\ll_\gamma, [E]^{\Gamma})$. Here we consider, as usual, the action law $\langle C, a, E\rangle$. We may recall that any information state γ, such that its projection $\mathcal{P}(\gamma)$ is the state w, can be taken as the point of reference of the employed ordering $\ll_\gamma$. For example, we can choose as γ the information state $\gamma(N(w))$ that corresponds to the hyper-neighbourhood $N(w)$. This results in the following reduction (for an action law $\langle C, a, E\rangle$ and an initial state w):

$$Res(w, a) = \{r \in \mathcal{D} : \ \mathcal{M}^*(\alpha, \beta) \text{ such that } \alpha \in min(\ll_{\gamma(N(w))}, [E]^{\Gamma}),$$
$$\beta \in \mathcal{K}_{\mathcal{M}}, \text{ and } \mathcal{P}(\beta) = r\}.$$

Informally, a successor state must satisfy the domain constraints D (belong to the set $\mathcal{D}$) and be a projection of some stable state that is $\mathcal{M}$-reachable from a state closest to $\gamma(N(w))$ among all states in $[E]^\Gamma$.

According to Lemma 7.2.2, the information state $\gamma(\|E\|_w)$ that corresponds to the trigger set, is the only $\ll_{\gamma(N(w))}$-minimal element:

$$\{\gamma(\|E\|_w)\} = min(\ll_{\gamma(N(w))}, [E]^\Gamma).$$

This fact allows us to simplify the selection function further:

$$Res(w,a) = \{r \in \mathcal{D} : \ \mathcal{M}^*(\gamma(\|E\|_w), \beta), \ \text{where} \ \beta \in \mathcal{K}_\mathcal{M} \ \text{and} \ \mathcal{P}(\beta) = r\}.$$

This selection function is identical to the selection function $Res_\Gamma(w,a)$ used in Chapter 5 to completely characterise the original selection function $Res_{RD\mathcal{L}}(w,a)$. Therefore, the construction described above allows us to obtain the first part of the following result.

Theorem 7.2.3 *For every action system based on causal relationships R, there exists a selection-equivalent action system $\langle \mathcal{W}, \ \mathcal{D}, \ \Gamma, \ \mathcal{E}, \ \mathcal{O}, \ \mathcal{M}, \ \Sigma \rangle$.*

Conversely, for every action system $\langle \mathcal{W}, \ \mathcal{D}, \ \Gamma, \ \mathcal{E}, \ \mathcal{O}, \ \mathcal{M}, \ \Sigma \rangle$, with the mini-choice choice function, there exists a selection-equivalent action system based on causal relationships R.

The second part of the theorem is essentially trivial if one parses the stratified binary relation $\mathcal{M}'$ into fully-qualified causal relationships. In other words, for every pair of information states such that $\mathcal{M}^*(\alpha, \beta)$ and $\beta \in \mathcal{K}_\mathcal{M}$ we consider the pair of states $x = \mathcal{P}(\alpha)$ and $y = \mathcal{P}(\beta)$. Then we create m^2 causal relationships

$$\epsilon_i \ \text{causes} \ \rho_j \ \text{if} \ \epsilon_1 \wedge \ldots \wedge \epsilon_i \wedge \ldots \wedge \epsilon_m,$$

where $x = \{\epsilon_1, \ldots, \epsilon_i, \ldots, \epsilon_m\}$ and $y = \{\rho_1, \ldots, \rho_j, \ldots, \rho_m\}$.

This selection-equivalence does not demand that an action system is Hamiltonian, compact or conservative. Of course, if needed, these characteristics could be added, resulting in a strengthening of the original approach of Thielscher [63]. For example, successor states satisfying the direct effects of actions are easily accommodated by conservative action systems, and compactness would be needed if the causal relationships R and the domain constraints D rule out exactly the same states.

7.2.4 Causal Propagation Semantics

Finally, we shall demonstrate selection-equivalence with Sandewall's action systems. This can be achieved with the state-space approximation $\mathcal{W} = \Gamma$ and $\mathcal{P}(\iota) = \iota$, staying within the same action domain $\mathcal{E}$ and the same set of admitted (legitimate) states $\mathcal{D}$.

To achieve the desired result, we take the "mini-choice gradient" $\sigma_M(r, e, w)$ as our choice function — one that chooses a $<_w$-minimal state in $[e] = [e]^\Gamma$. Also, we focus on conservative action systems. This results in the following reduction.

$$Res(w, e) = \{r \in \mathcal{D} \cap [e] : \ \mathcal{M}^*(\alpha, r), \text{ where } \alpha \in min(<_w, [e])\}.$$

Here, an admitted (and therefore, stable) state is a successor state if and only if it satisfies the direct effects of action e, and is, in addition, $\mathcal{M}$-reachable from a state closest to w among all states in $[e]$.

This construction together with setting $\mathcal{M} = C$ and the representation results of Chapter 6 allows us to achieve selection-equivalence based on $Res(w, e)$ and the original selection function:

$$Res_{C\mathcal{D}G}(w, e) = \{r_k \in \mathcal{D} : G(w, e, r_1), C(r_i, r_{i+1}), \lhd_w(r_i, r_{i+1}), 1 \leq i < k\}.$$

More precisely, it leads us to the first part of the following result, based on the representation lemmas 6.3.1 and 6.3.2 of Chapter 6.

Theorem 7.2.4 *For every respectful action system $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G\rangle$ there exists a selection-equivalent conservative action system $\langle \mathcal{W},\ \mathcal{D},\ \Gamma,\ \mathcal{E},\ \mathcal{O},\ \mathcal{M},\ \Sigma\rangle$, if the relation G satisfies conditions $(G_1) - (G_3)$.*

Conversely, for every conservative action system $\langle \mathcal{W},\ \mathcal{D},\ \Gamma,\ \mathcal{E},\ \mathcal{O},\ \mathcal{M}, \Sigma\rangle$, with $\mathcal{W} = \Gamma$ and the mini-choice choice function, there exists a selection-equivalent respectful action system $\langle \mathcal{W}, \mathcal{E}, C, \mathcal{D}, G\rangle$, if the orderings in $\mathcal{O}$ satisfy conditions $(M_1) - (M_5)$.

While the first part of the theorem uses $\mathcal{M} = C$, the second part needs an extraction of the causal relation C from a given relation $\mathcal{M}$ in a certain way. More precisely,

$$C(p, q) \quad \text{if and only if} \quad \mathcal{M}^*(p, q) \text{ and } q \in \mathcal{K}_\mathcal{M}.$$

In other words, $C = \mathcal{M}'$, where $\mathcal{M}'$ is a stratified relation extracted from $\mathcal{M}$ (as defined in Chapter 6) containing short-cuts to stable states. Together with conditions (M_4) and (M_5) Lemma 6.4.9 ensures that propagation along a transition chain starts from a $\prec_w$-minimal state, and links states that pair-wise respect the initial state. It should also be pointed out that if a respectful system is not required, then the sufficient conditions (M_4) and (M_5) can be dropped. This theorem establishes the conditions required to capture respectful action systems.

We may choose not to capture respectful action systems, in particular, but rather focus on conservative ones. After all, the most serious justification for respectful action systems was a requirement not to undo the direct effects of actions during causal propagation. Conservative action systems necessarily have successor states satisfying the direct effects of action e, while allowing the propagation to transit through states outside $[e]$, and sometimes "backtrack" in a direction opposite the PMA-ordering.

7.3 Classification and Discussion

In this section we intend to categorise variants of the general augmented preferential semantics according to the identified properties of action systems:

- conservatism, $Res(w, e) \subseteq [e]$,
- compactness, $\mathcal{D} = \mathcal{X}(\mathcal{K}_{\mathcal{M}})$,
- the Hamiltonian path condition ("extreme" context-sensitivity), and
- the approximation $\mathcal{W} = \Gamma$.

The latter property essentially allows us to characterise successor states without the information state-space. In addition, the classification scheme specifies the employed preferential structures and the appropriate choice functions:

	Causal fixed-points	Causal relationships	Causal propagation
Conservative	yes	no	yes
Compact	yes	no	no
Hamiltonian	yes	no	no
Information state-space	$\mathcal{W} = \Gamma$	$\mathcal{W} \neq \Gamma$	$\mathcal{W} \approx \Gamma$
Preferential structure	PMA	two-tiered PMA	PMA
Choice function	full-meet	mini-choice	mini-choice

Our characterisation of the causal propagation semantics was done under the approximation $\mathcal{W} = \Gamma$. The $\mathcal{W} \approx \Gamma$ classification entry for Sandewall's framework reflects a possibility to encode relevant action preconditions in information states, achieving the desired selection-equivalence without restricting the invocation relation G and orderings in $\mathcal{O}$.

In this chapter we considered the general augmented preferential semantics for causal reasoning about action. As mentioned above, the Principle of Minimal Change can be identified with the set of orderings, while causality is represented by the additional binary relation on states. Varying these and other components allows us to specify different instances of the framework, ranging from pure preferential semantics to causal propagation-oriented semantics.

In addition, the uniform framework of the general augmented preferential semantics allows us to easily compare existing approaches. In particular, the nature of the distinction between Sandewall's and Thielscher's approaches to propagation-oriented ramification may be uncovered and reduced to the variance in transition space dimensions and the employment of different preference metrics when identifying states nearest to the initial one.

In light of the general semantics, both of these cases (Sandewall's and Thielscher's) can be found very similar to McCain and Turner's causal theory of action using causal fixed-points [37]. This was also characterised by a variant of the augmented preferential semantics, relying on the PMA ordering and an appropriately constructed binary relation operating on the normal state-space. Here, the main difference appears to be the additional condition requiring that there exist a *Hamiltonian* path through certain states

in a state transition system, serving as a contextual mechanism and leading to causal fixed-points.

Essentially, the causal propagation employed by all our motivating approaches is context-sensitive. In order to capture various manifestations of context-sensitivity we identified a specific semantical component — a family of choice functions. In other words, rather than develop more sophisticated transition relations or preferential structures (that could have encoded context-sensitivity indirectly), we chose to capture this aspect explicitly. In fact, the appeal of an explicit representation of the underlying principles strongly motivated our overall approach, and is central to the general semantics presented in this work.

Chapter 8

Conclusion

8.1 Summary

In this dissertation we set out to examine the role of causality in reasoning about action and change, and develop a general semantics for a class of action theories. Our investigation covered a variety of semantics for causal reasoning about action and change — ranging from pure preferential semantics to its variants augmented with a causal relation. In our search we also attempted to highlight some of the aspects of "time's arrow" and causation relevant to an agent's reasoning process, while making the weakest possible semantical assumptions. In fact, we committed neither to a particular philosophic standpoint on the metaphysics of causation or minimality nor to any specific logical language. This approach allowed us to explore the role of several fundamental underlying principles transparently and objectively, on a set-theoretic basis, and without obscuring these principles by the internal structure of system components. The investigation culminated in a general unifying semantics for a broad class of causal action theories, represented by a number of recent influential approaches.

We focused on three of the most prominent causal frameworks in the Reasoning about Action literature: the causal systems with fixed-points suggested by McCain and Turner [37], the causal relationship approach of Thielscher [63], and Sandewall's causal propagation semantics [56] (the transition cascade semantics).

The reasons behind our expectations that these motivating approaches can be represented in a unifying setting were clarified in the first chapters of this dissertation. First of

all, we presented and analysed the multi-dimensional causation space (Figure 2.10) and placed our semantical framework in the ontological-epistemological plane of this space — this allowed us to characterise and contrast our motivating approaches systematically. Then, a simple variant of our framework for reasoning about actions and change was introduced in Chapter 2. This simple semantics describes causal reasoning about action as the propagation of an action's effects from minimal states (determined with respect to a chosen *preferential structure*) to stable states (ascertained with respect to some *causal relation*). Importantly, we identified the concept of the *information state-space*, where the propagation process takes place. This concept accentuates the differences between system dynamics (related to transitions between world states) and knowledge dynamics (involving transitions between information states). In addition, we established a valuable *density* condition linking legitimate system states and stable information states. The density condition ensures proper causal priority of the binary relation used in propagation.

As a next step, we studied possible areas of interaction between inertia and causality, in the context of action languages with different commitments towards causality and categorisation policies. Subsequently, we provided a preferential-style semantics (augmented with a causal relation on states) for the causal fixed-points framework of McCain and Turner and the causal relationship approach of Thielscher.

It is important to realise that, although the underlying constructions and proofs required to achieve the desired selection-equivalence are fairly complex, and the dimension of the information state-space Γ may be high (as is the case with the causal relationship approach), the semantics remains simple:

> *it describes, in simple terms, a process of propagation from the minimal elements (gradient area) to final state(s), followed by the projection onto the standard space $\mathcal{W}$.*

A variant of the augmented preferential semantics was also related to the causal propagation semantics (the transition cascade semantics) of Sandewall, subsuming it under certain uniformity assumptions. It is important to realise that, although the causal propagation semantics is sufficiently general and concise, it somewhat lacks the intuition behind the invocation relation G. If one is capable of specifying the precise immediate

effects of actions for each state of invocation, why is it not permitted to specify all final successor states directly as well (in other words, simply embed G in the selection function Res)? On the other hand, the preferential structure employed in our framework reflects the fundamental and intuitive Principle of Minimal Change, achieving the selection-equivalent identification of both the immediate and indirect effects of an action in the relevant state-space.

The additional assumptions stipulated in our characterisation of Sandewall's semantics are required if we wish to keep the direct effects of actions uniform — in other words, not dependent on the states of invocation. The price that we pay for it is the restrictions imposed on the invocation relation G in the causal propagation semantics, limiting the preconditions of an action. As mentioned earlier, in Chapter 6, instead of restricting the invocation relation G and orderings in $\mathcal{O}$, we may choose to abandon the approximation $\mathcal{W} = \Gamma$, and consider the information state-space Γ where relevant preconditions are properly encoded. In this case, the price is a higher dimension of the information state-space, and potentially a more convoluted projection function.

In summary, the unifying *general augmented preferential semantics*, emerging as a result of this study, captures two fundamental underlying principles — the *Principle of Minimal Change* and the *Principle of Causal Change* — and illustrates their clear and distinct roles.

Furthermore, the general semantics emphasises the role of contextual information affecting both minimality and causality and provides a means for balancing different contributing factors. In particular, by utilising *context-sensitive choice functions* such as full-meet gradient and mini-choice gradient, an agent scales and coordinates the causal propagation in the transition state-space. In other words, the general framework that we incrementally built in this work, allows an intelligent agent to represent a dynamic world in many ways, choosing certain components and discarding others, while staying within a uniform semantics.

It is argued that hidden or less immediate forces shaping our motivating approaches become transparent with the help of the general semantics. In addition, it is hoped that the unifying semantics would provide further insights into the views on causation and minimality, shared by these and other approaches.

8.2 Future Work

There are several interesting directions that can be pursued following the development of the general augmented preferential semantics.

One of them is an investigation into the applicability of the presented semantics to different approaches to Reasoning about Action. The instances considered in this work are already quite representative. However, there are a few other logics of action, and without a doubt, new proposals will appear in the near future. The immediate question is whether the semantics proposed here is general enough to deal with new theories of action, without any extensions. After all, it encapsulates a number of powerful concepts, such as the information state-space, preferential and causal structures, the family of choice functions. The potential variations of the components appear to be rich, and stretching the framework might be an interesting exercise. Even more importantly, identifying specific properties of action logics that place them outside of the proposed semantics may become a motivating research subject.

An additional challenge would be to verify which ones of the employed semantical components are most useful in characterising a given logic of action — in other words, investigate limits of applicability with various components. For example, it is quite clear that the causal relation and full-meet gradient, used in capturing the McCain and Turner theory, were the most principal components in that case, while the information state-space and the corresponding projection function played a significant role in characterising the causal relationship approach of Thielscher. Perhaps, future research will make some of the components marginal. The primary candidate for reductions would be the family of choice functions. One might attempt to encode causal context in the preferential structure and causal relation, already present in the semantics. A success in this direction will make the selection function more uniform using, for example, a mini-choice gradient on all occasions. We believe, however, that an explicit specification of some choice function (be it a mini-choice, a full-meet, or another function) helps in a better understanding of the causal context-sensitivity. Sometimes, the forces behind the causal propagation are more demanding than in other cases (compare, for instance, the full-meet gradient with the mini-choice one), and may require a precision that can be

more easily achieved with the dedicated component, rather than with a smart encoding. Nevertheless, this trade-off is an attractive topic for future work.

Another avenue for future work points toward a further comparison between the Principles of Minimal and Causal Change. Recent results reported in [43] indicated a simple way to distinguish between *minimisable* dynamic systems (that can be described by theories of action based on minimal change) and those systems that require *causal theories of action*, or in general, the systems "capable of forms of reasoning that cannot be captured by the Principle of Minimal Change" [43]. Their framework included a number of formal properties serving as necessary and sufficient conditions under which a dynamic system is minimisable. This establishes the range of applicability of the Principle of Minimal Change: "one simply needs to verify three properties" [43].

Interestingly, the McCain and Turner theory of causal fixed-points was shown to be minimisable, while the causal relationship approach of Thielscher was not. In particular, the causal fixed-points were characterised purely via a preferential structure, and without causal propagation — however, the employed preferential structure was defined over meta-states of a higher dimension. These meta-states played the role identical to information states in our semantics. In other words, the results presented in [43] described an alternative representation of the McCain and Turner theory in line with our semantics — this time, with the empty causal relation $\mathcal{M} = \emptyset$, but without the approximation $\mathcal{W} = \Gamma$. The following table allows us to compare different characterisations in terms of the employed preferential structures, the nature of causal propagation, and the possibility to approximate the information space Γ with the original state-space $\mathcal{W}$.

	Preferential structure	Propagation	Information space approximation
primary characterisation of McCain-Turner's theory	PMA $\prec_w$	$\mathcal{M} \neq \emptyset$	$\mathcal{W} = \Gamma$
alternative characterisation of McCain-Turner's theory	transitive ordering	$\mathcal{M} = \emptyset$	$\mathcal{W} \neq \Gamma$
characterisation of Thielscher's approach	two-tiered ordering $\ll_w$	$\mathcal{M} \neq \emptyset$	$\mathcal{W} \neq \Gamma$
restricted characterisation of Sandewall's semantics	$<_w$ satisfying $(M_1) - (M_5)$	$\mathcal{M} \neq \emptyset$	$\mathcal{W} = \Gamma$
unrestricted characterisation of Sandewall's semantics	transitive ordering	$\mathcal{M} \neq \emptyset$	$\mathcal{W} \neq \Gamma$

More importantly, the findings presented in [43] demonstrate that the general augmented preferential semantics may cover both minimisable and non-minimisable systems. The natural question is, whether this is too broad, and whether one would not be better off concentrating on non-minimisable (causal) systems independently from minimisable ones. In short, perhaps it is wiser to focus on distinctly causal theories of action and their unique properties. While this intention is definitely praiseworthy, we believe that the unifying framework presented in this dissertation provides an efficient way to compare minimisable and non-minimisable systems. For instance, in light of the results reported in [43], the systems with causal fixed-points and the systems based on causal relationships definitely differ with respect to being minimisable. However, the (alternative) representation of the former (based on $\mathcal{M} = \emptyset$ and $\mathcal{W} \neq \Gamma$) can be clearly differentiated now from the representation of the latter, that needed both $\mathcal{M} \neq \emptyset$ and $\mathcal{W} \neq \Gamma$. In addition, if one chooses our characterisation of causal fixed-points presented in Chapter 4, using $\mathcal{M} \neq \emptyset$ and $\mathcal{W} = \Gamma$, it becomes very clear that the approximation $\mathcal{W} = \Gamma$ has the information cost reflected in the *non-empty* causal relation $\mathcal{M}$. That is, there is a trade-off between these two components. These comments bring us closer to the subject of conciseness. As pointed out in [43],

> ... if the class of domains at focus is within the range of applicability of both causal and minimal change approaches, the determining factor in choosing between the two could be the "information cost" associated with the usage of each approach.

Developing a unifying semantics for a class of *concise* minimisable and non-minimisable (causal) logics of action appears at this stage to be one of the most appealing and engaging themes in Reasoning about Action and Change.

Appendix A

Proofs for Chapter 3

Lemma 3.3.2 *Let Ω^- be an $\mathcal{AC}^-$ domain description such that each non-inertial fluent is an effect-complete fluent:*

$$\mathcal{F} \setminus \Phi \subseteq \Sigma.$$

Then $\Omega_0 = \Omega^-$ defined in the signature $\langle\, \mathcal{F}, \emptyset, \mathcal{E} \,\rangle$ by abandoning the frame designation Φ is an $\mathcal{AC}_\mathcal{O}$ domain description, and the $\mathcal{AC}_\mathcal{O}$ models of Ω_0 are exactly the $\mathcal{AC}^-$ models of Ω^-.

Proof:

By construction $\Omega_0 = \Omega^-$. Consider now the set $Cn_\mathcal{Q}((s \cap s' \cap L_\Phi) \cup E)$. It is clear that

$$Cn_\mathcal{Q}((s \cap s' \cap L_\Phi) \cup E) \subseteq Cn_\mathcal{Q}((s \cap s') \cup E) \tag{A.1}$$

Let $L_\mathcal{F}$ be the set of all domain fluent literals, and L_Σ be the set of domain effect-complete fluent literals. Then

$$\begin{aligned}
&Cn_\mathcal{Q}((s \cap s') \cup E) = Cn_\mathcal{Q}((s \cap s' \cap L_\mathcal{F}) \cup E) = \\
&Cn_\mathcal{Q}((s \cap s' \cap (L_\Phi \cup (L_\mathcal{F} \setminus L_\Phi))) \cup E) = \\
&Cn_\mathcal{Q}((s \cap s' \cap L_\Phi) \cup (s \cap s' \cap (L_\mathcal{F} \setminus L_\Phi)) \cup E).
\end{aligned}$$

By the lemma assumption, $L_\mathcal{F} \setminus L_\Phi \subseteq L_\Sigma$. This entails

$$Cn_\mathcal{Q}((s \cap s') \cup E) \subseteq Cn_\mathcal{Q}((s \cap s' \cap L_\Phi) \cup (s \cap s' \cap L_\Sigma) \cup E).$$

Let us introduce the following abbreviations: $Y = s \cap s' \cap L_\Phi$ and $X = s \cap s' \cap L_\Sigma$. Then the right-hand side expression of the last containment can be abbreviated as $Cn_\mathcal{Q}(Y \cup X \cup E)$.

Given that every effect-complete literal in L_Σ is defined through inertial literals in L_Φ, it is easy to verify that for all φ, if $s \cap s' \cap L_\Sigma \mathrel{|\!\sim_\mathcal{Q}} \varphi$, then $s \cap s' \cap L_\Phi \mathrel{|\!\sim_\mathcal{Q}} \varphi$. In other words, if $\varphi \in Cn_\mathcal{Q}(s \cap s' \cap L_\Sigma)$ then $\varphi \in Cn_\mathcal{Q}(s \cap s' \cap L_\Phi)$. Hence,

$$Cn_\mathcal{Q}(X) \subseteq Cn_\mathcal{Q}(Y).$$

Therefore, for any set E, we obtain

$$Cn_\mathcal{Q}(X \cup Y \cup E) \subseteq Cn_\mathcal{Q}(Y \cup E).$$

Using this observation, we conclude that

$$Cn_\mathcal{Q}((s \cap s') \cup E) \subseteq Cn_\mathcal{Q}((s \cap s' \cap L_\Phi) \cup E) \tag{A.2}$$

Relationships A.1 and A.2 together establish that

$$Cn_\mathcal{Q}((s \cap s' \cap L_\Phi) \cup E) = Cn_\mathcal{Q}((s \cap s') \cup E)$$

Therefore, for every A and s, the sets of possible successor states are the same for the $\mathcal{AC}_\mathcal{O}$ and the $\mathcal{AC}^-$ state transitions:

$$Res_{\mathcal{AC}_\mathcal{O}}(s, A) = Res_{\mathcal{AC}^-}(s, A).$$

Consequently, the $\mathcal{AC}_\mathcal{O}$ models of Ω_0 are exactly the $\mathcal{AC}^-$ models of Ω^-.

■

Appendix B

Proofs for Chapter 4

Lemma 4.5.6 *For any two states r, w and sentence E,*

$$[(w \cap r) \cup \{E\}] = \langle\!\langle r, E \rangle\!\rangle_w.$$

Proof:

($\Longrightarrow$) We intend to prove that

$$[(w \cap r) \cup \{E\}] \subseteq \langle\!\langle r, E \rangle\!\rangle_w.$$

Let state $s \in [(w \cap r) \cup \{E\}]$, meaning in particular, $s \in [E]$. Let us assume that $s \notin \langle\!\langle r, E \rangle\!\rangle_w$. Then the state s is not a predecessor of r with respect to w, and therefore, *Diff*$(w, s) \not\subseteq$ *Diff*(w, r). Hence, there exists a literal h, such that $h \in$ *Diff*(w, s) and $h \notin$ *Diff*(w, r). Or, alternatively, $h \notin (w \cap s)$ and $h \in (w \cap r)$ — note that every state is a maximal consistent set of literals. Immediately, $h \notin s$ and $\neg h \in s$. Since $h \in (w \cap r)$, we obtain that

$$(w \cap r) \cup \{E\} \vdash \{h\}.$$

Also, $s \in [(w \cap r) \cup \{E\}]$. By definition of $[\cdot]$, state s is consistent with $(w \cap r) \cup \{E\}$. Since $\neg h \in s$, it follows that

$$\{\neg h\} \cup \lambda \vdash (w \cap r) \cup \{E\}$$

where $\lambda = s \setminus \{\neg h\}$ is a consistent subset of state s.

This leads to a contradiction $\{\neg h\} \cup \lambda \vdash \{h\}$, showing that $s \in \langle\!\langle r, E \rangle\!\rangle_w$, and hence,

$$[(w \cap r) \cup \{E\}] \subseteq \langle\!\langle r, E \rangle\!\rangle_w.$$

($\Longleftarrow$) Now we need to prove that

$$\langle\!\langle r, E \rangle\!\rangle_w \subseteq [(w \cap r) \cup \{E\}].$$

Let $s \in \langle\!\langle r, E \rangle\!\rangle_w$. Then $s \in [E]$, and moreover, ***Diff***$(w, s) \subseteq$ ***Diff***(w, r). Consequently, $(w \cap r) \subseteq (w \cap s)$, and $[w \cap s] \subseteq [w \cap r]$. Hence,

$$[(w \cap s) \cup \{E\}] \subseteq [(w \cap r) \cup \{E\}].$$

Since $s \in [E]$, we obtain $s \in [(w \cap s) \cup \{E\}]$, and therefore, $s \in [(w \cap r) \cup \{E\}]$. This establishes that

$$\langle\!\langle r, E \rangle\!\rangle_w \subseteq [(w \cap r) \cup \{E\}].$$

■

Theorem 4.5.7 *For every causal system there exists a selection-equivalent state elimination system. Conversely, for every state elimination system there exists a selection-equivalent causal system.*

Proof:

($\Longrightarrow$) Let $\mathcal{Q}$ be an arbitrary causal system. For every causal rule $\varphi \Rightarrow \psi$ in $\mathcal{Q}$, produce the elimination rule $[\varphi] \rhd [\varphi \wedge \psi]$. Call $\mathcal{S}$ the set of elimination rules so produced. We shall verify that for any legitimate state w and sentence E,

$$Res_{\mathcal{Q}}(w,\ E) = \mathtt{Next}_{\mathcal{S}}(w,\ E).$$

(a) Let state $r \in Res_{\mathcal{Q}}(w,\ E)$. We need to show $r \in \mathtt{Next}_{\mathcal{S}}(w,\ E)$. The state r is a causal fixed-point. By definition of $Res_{\mathcal{Q}}(w,\ E)$, state r is consistent with $(w \cap r) \cup \{E\}$. Then by definition of $[\cdot]$, $r \in [(w \cap r) \cup \{E\}]$.

The state elimination system $\mathcal{S}$ includes an elimination rule $[\varphi_i] \rhd [\varphi_i \wedge \psi_i]$ for every causal rule $\varphi_i \Rightarrow \psi_i$ in the causal system $\mathcal{Q}$. Let

$$[\varphi_1] \rhd [\varphi_1 \wedge \psi_1]$$

be a rule such that

$$[(w \cap r) \cup \{E\}] \subseteq [\varphi_1].$$

By definition of $\rightsquigarrow$,

$$[(w \cap r) \cup \{E\}] \ \rightsquigarrow \ [(w \cap r) \cup \{E\}] \cap [\varphi_1 \wedge \psi_1].$$

Let us now consider a sequence of causal rules $\varphi_i \Rightarrow \psi_i$ ($1 \leq i < n$), such that their successive application eliminates as many states in $[(w \cap r) \cup \{E\}]$ as possible. In other words, for each $i, 1 \leq i < n$, we require

$$[(w \cap r) \cup \{E\}] \cap [\varphi_1 \wedge \psi_1] \cap \ldots \cap [\varphi_i \wedge \psi_i] \subseteq [\varphi_{i+1}].$$

This process results in

$$\begin{aligned}
[(w \cap r) \cup \{E\}] \ &\rightsquigarrow \ [(w \cap r) \cup \{E\}] \cap [\varphi_1 \wedge \psi_1] \\
&\rightsquigarrow \ [(w \cap r) \cup \{E\}] \cap [\varphi_1 \wedge \psi_1] \cap [\varphi_2 \wedge \psi_2] \ \rightsquigarrow \ \ldots \\
&\rightsquigarrow \ [(w \cap r) \cup \{E\}] \cap [\varphi_1 \wedge \psi_1] \cap \ldots \cap [\varphi_i \wedge \psi_i] \ \rightsquigarrow \ \ldots \\
&\rightsquigarrow \ [(w \cap r) \cup \{E\}] \cap [\varphi_1 \wedge \psi_1] \cap \ldots \cap [\varphi_n \wedge \psi_n],
\end{aligned}$$

where n is the number of all applicable elimination rules used in filtering of the set $[(w \cap r) \cup \{E\}]$.

We focus now on the last right-hand side expression

$$[(w \cap r) \cup \{E\}] \cap [\varphi_1 \wedge \psi_1] \cap \ldots \cap [\varphi_n \wedge \psi_n],$$

and intend to show that the state r is the only element of this set:

$$\{r\} = [(w \cap r) \cup \{E\}] \cap [\varphi_1 \wedge \psi_1] \cap \ldots \cap [\varphi_n \wedge \psi_n].$$

First of all, the state r is a causal fixed-point and must satisfy all the causal rules. Then it would belong to the intersection $[\varphi_1 \wedge \psi_1] \cap \ldots \cap [\varphi_n \wedge \psi_n]$. Also, the Lemma 4.5.6 established that

$$[(w \cap r) \cup \{E\}] = \langle\!\langle r, E \rangle\!\rangle_w.$$

Since $r \in \langle\!\langle r, E \rangle\!\rangle_w$, we obtain $r \in [(w \cap r) \cup \{E\}]$. Consequently,

$$r \in [(w \cap r) \cup \{E\}] \cap [\varphi_1 \wedge \psi_1] \cap \ldots \cap [\varphi_n \wedge \psi_n].$$

Now we need to show that there is no other state r' in the set on the right-hand side. Assume the opposite: there exists a state $r' \neq r$, such that

$$r' \in [(w \cap r) \cup \{E\}] \cap [\varphi_1 \wedge \psi_1] \cap \ldots \cap [\varphi_n \wedge \psi_n].$$

The assumption $r' \in [(w \cap r) \cup \{E\}] = \langle\!\langle r, E \rangle\!\rangle_w$ means that the state r' is at least as close to w as the state r in the PMA ordering. Moreover, since $r' \neq r$, the state r' is *closer* to w than the state r. In other words, there is at least one literal $h \in (w \cap r')$ such that $h \notin (w \cap r)$. It follows that $\neg h \in r$. This would mean that there was a causal rule $\varphi_i \Rightarrow \psi_i$, where $\neg h \in [\varphi_i \wedge \psi_i]$. The latter conclusion contradicts our assumption that $r' \in [\varphi_1 \wedge \psi_1] \cap \ldots \cap [\varphi_n \wedge \psi_n]$. In other words, the "extra" literal $\neg h$ might have appeared in the state r only as a result of causal inference, and then this inference cannot support the literal h.

Therefore, r is the only element of the considered set, and the elimination process yields

$$[(w \cap r) \cup \{E\}] \;\overset{*}{\rightsquigarrow}\; \{r\},$$

where state r is a final state in $\mathcal{S}$. Using the Lemma 4.5.6 again, we obtain

$$\langle\!\langle r, E \rangle\!\rangle_w \;\overset{*}{\rightsquigarrow}\; r,$$

meaning that $r \in \mathtt{Next}_{\mathcal{S}}(w,\ E)$, and hence

$$Res_{\mathcal{Q}}(w,\ E) \subseteq \mathtt{Next}_{\mathcal{S}}(w,\ E).$$

(b) Now let state $r \in \mathtt{Next}_{\mathcal{S}}(w,\ E)$. We need to show $r \in Res_{\mathcal{Q}}(w,\ E)$. Starting with

$$\langle\!\langle r, E \rangle\!\rangle_w \;\overset{*}{\rightsquigarrow}\; r$$

and using the Lemma 4.5.6, we obtain

$$[(w \cap r) \cup \{E\}] \;\overset{*}{\rightsquigarrow}\; r,$$

where r is a final state in $\mathcal{S}$.

Reversing the elimination process described above, we conclude that the state r satisfies all causal rules $\varphi_i \Rightarrow \psi_i$ in the causal system $\mathcal{Q}$, applicable to $[(w \cap r) \cup \{E\}]$.

Therefore, it is a causal fixed-point, $r \in Res_{\mathcal{Q}}(w,\ E)$, verifying that for any legitimate state w and sentence E,

$$Res_{\mathcal{Q}}(w,\ E) = \mathtt{Next}_{\mathcal{S}}(w,\ E).$$

($\Longleftarrow$) Let $\mathcal{S}$ be an arbitrary state elimination system. For every elimination rule $X \triangleright Y$ produce the causal law $\varphi \Rightarrow \psi$, where $\varphi,\ \psi$ are such that $[\varphi] = X$ and $[\psi] = Y$ (since our language is a finitary propositional one, such φ and ψ always exist). We intend to prove that the set of causal laws so produced, call it $\mathcal{Q}$, is selection-equivalent to $\mathcal{S}$:

$$\mathtt{Next}_{\mathcal{S}}(w,\ E) = Res_{\mathcal{Q}}(w,\ E).$$

(a) Let state $r \in \mathtt{Next}_{\mathcal{S}}(w,\ E)$. We need to show $r \in Res_{\mathcal{Q}}(w,\ E)$. As in the previous case we start with

$$\langle\!\langle r, E \rangle\!\rangle_w \ \overset{*}{\rightsquigarrow}\ r$$

and use the Lemma 4.5.6, in obtaining

$$[(w \cap r) \cup \{E\}] \ \overset{*}{\rightsquigarrow}\ r,$$

where r is a final state in $\mathcal{S}$. Now, given the "engineered" nature of the causal system $\mathcal{Q}$, it is easy to see that there is a chain of elimination rules

$$[(w \cap r) \cup \{E\}] = X_1 \triangleright X_2 \triangleright\ \ldots \triangleright X_{n-1} \triangleright X_n = \{r\},$$

where for each i ($1 < i \leq n$), $X_i \subseteq X_{i-1}$, and a chain of causal rules

$$(w \cap r) \cup \{E\} = \varphi_1 \Rightarrow \varphi_2 \Rightarrow\ \ldots \Rightarrow \varphi_{n-1} \Rightarrow \varphi_n = \bigwedge r,$$

where $\bigwedge r$ is a conjunction of all literals in the state r. In other words, every literal in r can be causally inferred from the sentence $(w \cap r) \cup \{E\}$, meaning that r is a causal fixed-point, $r \in Res_{\mathcal{Q}}(w,\ E)$. Therefore,

$$\mathtt{Next}_{\mathcal{S}}(w,\ E) \subseteq Res_{\mathcal{Q}}(w,\ E).$$

(b) Now let state $r \in Res_{\mathcal{Q}}(w,\ E)$. We need to show $r \in \mathtt{Next}_{\mathcal{S}}(w,\ E)$. By definition of $Res_{\mathcal{Q}}(w,\ E)$, state r is consistent with $(w \cap r) \cup \{E\}$. Then by definition of $[\cdot]$, $r \in [(w \cap r) \cup \{E\}]$.

Repeating the elimination process from the earlier proof (a) of the ($\Longrightarrow$) part, we obtain[1]

$$[(w \cap r) \cup \{E\}] \ \stackrel{*}{\rightsquigarrow} \ r,$$

where state r is a final state in $\mathcal{S}$. Using the Lemma 4.5.6, we obtain

$$\langle\!\langle r, E \rangle\!\rangle_w \ \stackrel{*}{\rightsquigarrow} \ r,$$

meaning that $r \in \mathtt{Next}_{\mathcal{S}}(w,\ E)$, and hence

$$Res_{\mathcal{Q}}(w,\ E) \subseteq \mathtt{Next}_{\mathcal{S}}(w,\ E).$$

This concludes the proof, verifying that for any legitimate state w and sentence E,

$$\mathtt{Next}_{\mathcal{S}}(w,\ E) = Res_{\mathcal{Q}}(w,\ E).$$

■

Theorem 4.6.5 *For every state elimination system $\mathcal{S}$ there is a selection-equivalent state transition system $\mathcal{M}$. Conversely, for every state transition system $\mathcal{M}$ there is a selection-equivalent state elimination system $\mathcal{S}$.*

Proof:

($\Longrightarrow$) Let $\mathcal{S}$ be a state elimination system. Let $\mathcal{S}'$ be a unary state elimination system that is selection-equivalent to $\mathcal{S}$. From $\mathcal{S}'$ we construct a selection-equivalent state transition system $\mathcal{M}$ in the following manner.

For any two states r and r', we shall specify $\mathcal{M}(r, r')$ if and only if there is a dissolvable set of states Π containing r and r', such that for some trace of Π in $\mathcal{S}'$, r' appears immediately after r.

We intend to show that $\mathcal{M}$ is selection-equivalent to $\mathcal{S}'$:

$$\mathtt{Next}_{\mathcal{S}'}(w,\ E) = \mathtt{Succ}_{\mathcal{M}}(w,\ E).$$

(a) Let state $r \in \mathtt{Next}_{\mathcal{S}'}(w,\ E)$. We need to show $r \in \mathtt{Succ}_{\mathcal{M}}(w,\ E)$.

[1] Please note that now we use the causal rules $\varphi \Rightarrow \psi$ constructed from the elimination rules $X \triangleright Y$ as follows: $[\varphi] = X$ and $[\psi] = Y$.

First of all, since $r \in \texttt{Next}_{\mathcal{S}'}(w,\ E)$, the state $r \in [E]$. Also, it is not difficult to see that r is final in $\mathcal{M}$. If it was no so, there would exist a state q such that $\mathcal{M}(r, q)$. This, in turn, would require an existence of a trace

$$r_1; r_2; \cdots; r; q; \cdots;$$

and a corresponding string of elimination rules

$$\sigma_1; \sigma_2; \cdots; \sigma_r; \cdots,$$

where the rule σ_r is the rule $\{r, q, \ldots\} \rhd \{q, \ldots\}$ eliminating state r. More precisely, it can be derived from the rule σ_r that $\{r\} \rightsquigarrow \emptyset$, contradicting with r being final in $\mathcal{S}'$. Hence, r is final in $\mathcal{M}$, and the only remaining part to show here is that there exists a Hamiltonian path through states in $\langle\!\langle r, E \rangle\!\rangle_w$.

Let us assume, without loss of generality, that the set $\langle\!\langle r, E \rangle\!\rangle_w$ has n elements ($n \geq 1$). Since $\langle\!\langle r, E \rangle\!\rangle_w \overset{*}{\rightsquigarrow} r$ (because r is an element of $\texttt{Next}_{\mathcal{S}'}(w,\ E)$), there is a sequence of elimination rules

$$\sigma_1; \cdots; \sigma_i; \cdots; \sigma_{n-1}$$

such that

$$\sigma_i \text{ is } X_i \rhd X_{i+1},$$

where X_1 is chosen as $\langle\!\langle r, E \rangle\!\rangle_w$, $X_{n-1} = \{r_{n-1}, r\}$ for some state r_{n-1}, $X_n = \{r\}$, and each $X_i \setminus X_{i+1}$ is a singleton, while $X_{i+1} \subset X_i$ ($1 \leq i < n$). Then

$$X_1 \setminus X_2; \cdots; X_i \setminus X_{i+1}; \cdots; X_{n-1} \setminus X_n; X_n$$

is a trace, equivalently represented as (denoting $X_i \setminus X_{i+1}$ by r_i)

$$r_1; \cdots; r_i; \cdots; r_{n-1}; r_n = r,$$

where the set $\langle\!\langle r, E \rangle\!\rangle_w$ is a union of all r_i ($1 \leq i \leq n$). Therefore, $\mathcal{M}(r_i, r_{i+1})$ for all i, $1 \leq i < n$. This effectively constructs a Hamiltonian path through the states in $\langle\!\langle r, E \rangle\!\rangle_w$. Hence, $r \in \texttt{Succ}_{\mathcal{M}}(w,\ E)$, and

$$\texttt{Next}_{\mathcal{S}'}(w,\ E) \subseteq \texttt{Succ}_{\mathcal{M}}(w,\ E).$$

(b) Now let state $r \in \texttt{Succ}_{\mathcal{M}}(w,\ E)$. We need to show $r \in \texttt{Next}_{\mathcal{S}'}(w,\ E)$. Since $r \in \texttt{Succ}_{\mathcal{M}}(w,\ E)$, the state $r \in [E]$. It is also clear that r is final in $\mathcal{S}'$. If this was not so, there would exist a state q such that $r \rightsquigarrow q$. This would presuppose an elimination rule $\{r, q, \delta\} \rhd \{q, \delta\}$, where δ is a sequence of states such that $r \notin \delta$. Assuming existence of at least one state $p \in \delta$ final in $\mathcal{S}'$ (otherwise, if there are no final states and no traces, the proof is trivial), we obtain a trace $r; q; \ldots; p$, resulting in $\mathcal{M}(r, q)$. This contradicts r being a final state in $\mathcal{M}$. Hence, r is final in $\mathcal{S}'$, and we only need to show now that $\langle\!|r, E|\!\rangle_w \overset{*}{\rightsquigarrow} \{r\}$.

Since $r \in \texttt{Succ}_{\mathcal{M}}(w,\ E)$, there exists a Hamiltonian path through states in $\langle\!|r, E|\!\rangle_w$. Let us assume that $\langle\!|r, E|\!\rangle_w = \{r_1, \ldots, r_i, \ldots, r_n\}$, and let the sequence of states

$$r_1; \cdots; r_i; \cdots; r_{n-1}; r_n = r,$$

be such a path. This sequence is, by Definition 4.6.4 and construction of $\mathcal{M}$, a trace. Hence, there exists a string of elimination rules

$$\sigma_1; \cdots; \sigma_i; \cdots; \sigma_{n-1}$$

such that for all i ($1 \leq i < n$), the rule $\sigma_i = \{r_i, r_{i+1}, \ldots, r_n\} \rhd \{r_{i+1}, \ldots, r_n\}$ eliminates the state r_i. Therefore,

$$\langle\!|r, E|\!\rangle_w = \{r_1, \ldots, r_i, \ldots, r_n\} \overset{*}{\rightsquigarrow} \{r_n\},$$

and consequently $r \in \texttt{Next}_{\mathcal{S}'}(w,\ E)$. Hence,

$$\texttt{Succ}_{\mathcal{M}}(w,\ E) \subseteq \texttt{Next}_{\mathcal{S}'}(w,\ E).$$

This concludes the proof, verifying that for any legitimate state w and sentence E,

$$\texttt{Next}_{\mathcal{S}'}(w,\ E) = \texttt{Succ}_{\mathcal{M}}(w,\ E).$$

($\Longleftarrow$) Let $\mathcal{M}$ be a state transition system. From $\mathcal{M}$ we construct a selection-equivalent unary state elimination system $\mathcal{S}'$ in the following manner.

For any two states r and r', we shall specify $\{r, r'\} \rhd \{r'\}$ if and only if $\mathcal{M}(r, r')$.

Given that the state elimination system $\mathcal{S}'$ is "reverse-engineered" from the construction presented above, the proofs are identical to those of ($\Longrightarrow$), and show that $\mathcal{S}'$ is selection-equivalent to $\mathcal{M}$:

$$\texttt{Succ}_{\mathcal{M}}(w,\ E) = \texttt{Next}_{\mathcal{S}'}(w,\ E).$$

■

Appendix C

Proofs for Chapter 5

Lemma 5.1.4 *If* $(s', E') \overset{*}{\rightsquigarrow} (s'', E'')$*, then* $E'' \subseteq s''$.

Proof:

First of all, we observe that any initial pair (s, E) is constructed as $((w \setminus C) \cup E, E)$, where w is the initial state and C, E are the action condition and effect respectively. Clearly, if a literal ϵ is in E, then it belongs to $s = (w \setminus C) \cup E$ as well. Hence,

$$E \subseteq s. \tag{C.1}$$

Furthermore, by definition 5.1.2, if $(s, E) \rightsquigarrow (s', E')$, then

$$s' = (s \setminus \{\neg\rho\}) \cup \{\rho\} \tag{C.2}$$

$$E' = (E \setminus \{\neg\rho\}) \cup \{\rho\}. \tag{C.3}$$

In other words, both propagated components - current state and current effects - are updated with respect to the effect ρ simultaneously and analogously. Using C.1, C.2 and C.3, we obtain $E' \subseteq s'$. A simple induction on a length of a $(s_0, E_0) \rightsquigarrow (s_1, E_1) \rightsquigarrow ... \rightsquigarrow (s_n, E_n)$ shows that $E_n \subseteq s_n$. Therefore, the property $E' \subseteq s'$ holds for every pair (s', E') if $(s, E) \overset{*}{\rightsquigarrow} (s', E')$.

■

Lemma 5.2.1 *If* $f \in \mathcal{F}$*, then* $l(\neg f) = \neg l(f)$.

Proof:

To prove it, we note that, if $f \in \mathcal{F}$, then $\neg f$ is a negative literal. By definition of l and $|f|$, $l(\neg f) = \neg j(|f|) = \neg j(f) = \neg l(f)$. ■

Lemma 5.4.4 *For any two states $x \in \mathcal{W}$ and $y \in \mathcal{W}$, if the connection set $L(x,y) \neq \emptyset$ then there exists a justifier literal $\overset{\circ}{f}$ such that $[\overset{\circ}{f}] \cap N(x) \subseteq L(x,y)$.*

Proof:

Let $x \in \mathcal{W}$ and $y \in \mathcal{W}$ be two states. Suppose $L(x,y) \neq \emptyset$. We need to show that there exists a literal $\overset{\circ}{f}$ such that $[\overset{\circ}{f}] \cap N(x) \subseteq L(x,y)$.

Since $L(x,y) \neq \emptyset$, then by definition of $L(x,y)$ there exists a $\mathcal{C}$-link between a hyper-state in $N(x)$ and a hyper state in $N(y)$. That is, $\mathcal{C}(s_1, s_2)$ for $s_1 \in N(x)$ and $s_2 \in N(y)$. Then $p(s_1) = x$ and $p(s_2) = y$ by definition 5.2.4.

By definition 5.2.6, there exists a causal relationship ϵ causes ρ if Φ, such that

$$p(s_1) \vdash \epsilon \wedge \Phi \wedge \neg\rho \tag{C.4}$$

$$h(s_1) \vdash \overset{\circ}{\epsilon} \tag{C.5}$$

$$p(s_2) = (p(s_1) \setminus \{\neg\rho\}) \cup \{\rho\} \tag{C.6}$$

$$h(s_2) = (h(s_1) \setminus \{\neg\overset{\circ}{\rho}\}) \cup \{\overset{\circ}{\rho}\} \tag{C.7}$$

Consider the set $B_\epsilon = [\overset{\circ}{\epsilon}] \cap N(x)$ and an element $s \in B_\epsilon$. We need to show that $s \in L(x,y)$.

For any state $s \in B_\epsilon$ we have $s \in N(x)$ and $s \in [\overset{\circ}{\epsilon}]$, which means

$$h(s) \vdash \overset{\circ}{\epsilon}. \tag{C.8}$$

States s, s_1 belong to the same neighbourhood $N(x)$, and therefore, $p(s) = x = p(s_1)$ by definition 5.2.4. Hence, using C.4, we obtain that

$$p(s) \vdash \epsilon \wedge \Phi \wedge \neg\rho. \tag{C.9}$$

Now consider the set $s' = (s \setminus \{\neg\rho, \neg\overset{\circ}{\rho}\}) \cup \{\rho, \overset{\circ}{\rho}\}$ (obviously, s' is a state). Clearly,

$$p(s') = (p(s) \setminus \{\neg\rho\}) \cup \{\rho\} \tag{C.10}$$

$$h(s') = (h(s) \setminus \{\neg\overset{\circ}{\rho}\}) \cup \{\overset{\circ}{\rho}\}. \tag{C.11}$$

Putting C.8, C.9, C.10, C.11 together and using the definition 5.2.6, we obtain that $\mathcal{C}(s, s')$ holds. From C.6 and the fact that $p(s) = p(s_1)$ it follows that $p(s_2) = (p(s) \setminus$

$\{\neg\rho\}) \cup \{\rho\}$. Hence, using C.10, we obtain $p(s_2) = p(s') = y$. Therefore, $s' \in N(y)$. By definition of $L(x, y)$ it follows that $s \in L(x, y)$. Therefore, $\overset{\circ}{\epsilon}$ is a suitable candidate and $[\overset{\circ}{\epsilon}] \cap N(x) \subseteq L(x, y)$ as desired.

■

Lemma 5.4.5 *For any two states $x \in \mathcal{W}$ and $y \in \mathcal{W}$, if there exists a justifier literal $\overset{\circ}{f}$ such that $[\overset{\circ}{f}] \cap N(x) \subseteq L(x, y)$, then there exists a causal relationship f* causes *ρ* if *Φ, for some Φ true in x, where $\{\rho\} = y \setminus x$.*

Proof:

Let $x \in \mathcal{W}$ and $y \in \mathcal{W}$ be two states. Suppose that there exists a justifier literal $\overset{\circ}{f}$ such that $[\overset{\circ}{f}] \cap N(x) \subseteq L(x, y)$. We need to show that there exists a causal relationship f causes ρ if Φ, for some Φ and where $\{\rho\} = y \setminus x$.

Consider the set $B_f = [\overset{\circ}{f}] \cap N(x)$ and its element - a state $s_i \in B_f$. We have then $s_i \in N(x)$ which yields, by definition 5.2.4,

$$p(s_i) = x. \tag{C.12}$$

By lemma assumption $s_i \in L(x, y)$, and hence $\mathcal{C}(s_i, s_i')$ holds for some $s_i' \in N(y)$. By definition 5.2.6, there exists a causal relationship $cr_i : \epsilon_i$ causes ρ if Φ_i, where

$$p(s_i) \vdash \epsilon_i \wedge \Phi_i \wedge \neg\rho \tag{C.13}$$

$$p(s_i') = (p(s_i) \setminus \{\neg\rho\}) \cup \{\rho\} \tag{C.14}$$

for some Φ_i. Using C.13 and C.14, we obtain that $\{\rho\} = p(s_i') \setminus p(s_i)$. From C.12 and the fact that $s_i' \in N(y)$ (or, equivalently, $p(s_i') = y$), it follows that

$$\{\rho\} = y \setminus x. \tag{C.15}$$

Using C.12, C.13 and setting $\Phi = \Phi_i$, we obtain that

$$x \vdash \Phi. \tag{C.16}$$

We only need to show now that $f = \epsilon_i$ for some i.

Let us assume that this is not the case, and for all ϵ_i, $\epsilon_i \neq f$. In other words, all possible causal relationships cr_i have causes ϵ_i distinct from the literal f. Using C.13,

we can see that $p(s_i) \vdash \epsilon_i$. Using C.12 we observe that $x = p(s_i)$ for all i, and therefore, due to consistency of the state $x \in \mathcal{W}$, there are no two causal relationships cr_i and cr_j among ones under consideration such that $\epsilon_i = \neg\epsilon_j$. Hence, there are at most $m-1$ different justifier literals $\overset{\circ}{\epsilon_i}$ (as the literal f is excluded). It is easy to show that varying n ($n \leq m$) distinct justifier literals $\overset{\circ}{\epsilon_i} \neq \neg\overset{\circ}{\epsilon_j}$ (and keeping $m-n$ justifier literals fixed), accounts for precisely $\Sigma_{k=m-n+1}^{m} 2^{m-k}$ states in any hyper-neighbourhood $N(x)$. Hence, varying $m-1$ justifier literals $\overset{\circ}{\epsilon_i}$ accounts for at most $\Sigma_{k=m-(m-1)+1}^{m} 2^{m-k} = 2^{m-1} - 1$ states. However, there are 2^{m-1} states in the set B_f, which were obtained by fixing the justifier literal $\overset{\circ}{f}$.

The contradiction shows that $f = \epsilon_i$ for some i, and therefore, there exists a causal relationship f causes ρ if Φ, for some Φ and ρ, where, as shown by C.15 and C.16, $\{\rho\} = y \setminus x$, and $x \vdash \Phi$.

■

Lemma 5.4.7 *For any two states $x \in \mathcal{W}$ and $y \in \mathcal{W}$, there is no justifier literal $\overset{\circ}{\epsilon}$ such that both $[\overset{\circ}{\epsilon}] \cap N(x) \subseteq L(x,y)$ and $[\neg\overset{\circ}{\epsilon}] \cap N(x) \subseteq L(x,y)$ hold.*

Proof:

Let $x \in \mathcal{W}$ and $y \in \mathcal{W}$ be two states. We need to show that there is no justifier literal $\overset{\circ}{\epsilon}$ such that both $[\overset{\circ}{\epsilon}] \cap N(x) \subseteq L(x,y)$ and $[\neg\overset{\circ}{\epsilon}] \cap N(x) \subseteq L(x,y)$ hold.

If $L(x,y) = \emptyset$, the result follows trivially. Suppose $L(x,y) \neq \emptyset$. By lemma 5.4.4 there exists a justifier literal $\overset{\circ}{\epsilon}$ such that

$$[\overset{\circ}{\epsilon}] \cap N(x) \subseteq L(x,y). \tag{C.17}$$

We now need to show that $[\neg\overset{\circ}{\epsilon}] \cap N(x) \subseteq L(x,y)$ does not hold. To do so we assume the opposite:

$$[\neg\overset{\circ}{\epsilon}] \cap N(x) \subseteq L(x,y). \tag{C.18}$$

Then the corollary 5.4.6 and C.17, C.18 together yield that there exist two causal relationships $\overset{\circ}{\epsilon}$ causes ρ if Φ_1 and $\neg\overset{\circ}{\epsilon}$ causes ρ if Φ_2 for some Φ_1 and Φ_2. Therefore, by definition 5.2.6 there exist two different C-links $C(s_1, s_1')$ and $C(s_2, s_2')$, where $s_1 \in N(x)$ and $s_2 \in N(x)$ such that

1. $p(s_1) \vdash \epsilon \wedge \Phi_1 \wedge \neg\rho$

2. $p(s_2) \vdash \neg\epsilon \wedge \Phi_2 \wedge \neg\rho$

It clearly follows that

$$p(s_1) \vdash \epsilon \tag{C.19}$$

$$p(s_2) \vdash \neg\epsilon. \tag{C.20}$$

Using $s_1 \in N(x)$ and $s_2 \in N(x)$ and the definition 5.2.4, we obtain that $p(s_1) = p(s_2) = x$. Since the state $x \in \mathcal{W}$ is consistent, the latter observation contradicts C.19 and C.20.

■

Lemma 5.4.9 *For any initial state $w \in \mathcal{W}$ and an action a, where $\langle C, a, E\rangle$ is the action law, $\cap_{s\in\|E\|_w} h(s) = \overset{\circ}{E}$.*

Proof:

Let E be effects of some action law $< C, a, E >$ and let $\|E\|_w$ be the trigger set defined by 5.4.8. We need to show that $\cap_{s\in\|E\|_w} h(s) = \overset{\circ}{E}$.

We first show that $\overset{\circ}{E} \subseteq \cap_{s\in\|E\|_w} h(s)$. Consider an element of $\overset{\circ}{E}$. If $\overset{\circ}{f} \in \overset{\circ}{E}$, then, by definition 5.4.8 of $\|E\|_w$, $h(s) \vdash \overset{\circ}{f}$ for all states $s \in \|E\|_w$. Hence, $\overset{\circ}{f} \in \cap_{s\in\|E\|_w} h(s)$, and $\overset{\circ}{E} \subseteq \cap_{s\in\|E\|_w} h(s)$.

In order to show the reverse part of the containment, $\cap_{s\in\|E\|_w} h(s) \subseteq \overset{\circ}{E}$, we consider an element $\overset{\circ}{f} \in \cap_{s\in\|E\|_w} h(s)$. Clearly,

$$h(s) \vdash \overset{\circ}{f} \text{ for all states } s \in \|E\|_w. \tag{C.21}$$

We need to show that $\overset{\circ}{f} \in \overset{\circ}{E}$. Let us assume the opposite: $\overset{\circ}{f} \notin \overset{\circ}{E}$. Consider the set $\overset{\circ}{U} = \overset{\circ}{E} \cup \{\overset{\circ}{f}\}$. By definition 5.4.8 of $\|E\|_w$,

$$h(s) \vdash \overset{\circ}{E} \text{ for all states } s \in \|E\|_w. \tag{C.22}$$

Combining C.21 and C.22, we obtain $h(s) \vdash \overset{\circ}{U}$ for all states $s \in \|E\|_w$. Then, by definition 5.4.8, $\|U\|_w$ is the trigger set, and $\overset{\circ}{U} = \overset{\circ}{E}$, leading to a contradiction with $\overset{\circ}{f} \notin \overset{\circ}{E}$. Therefore, $\overset{\circ}{f} \in \overset{\circ}{E}$, and $\cap_{s\in\|E\|_w} h(s) \subseteq \overset{\circ}{E}$.

Thus, $\cap_{s\in\|E\|_w} h(s) = \overset{\circ}{E}$.

■

Lemma 5.4.13 *If $\|E\|_w \subset N(x)$, then $\|E\|_w \succ N(y)$ for some $y \in \mathcal{W}$ if and only if $(x, E) \overset{*}{\rightsquigarrow} (y, E')$ for some E'.*

Proof:

($\Longrightarrow$) Let $\|E\|_w \subseteq N(x)$, and $\|E\|_w \succ N(y)$ for some $y \in \mathcal{W}$. We need to show that $(x, E) \overset{*}{\rightsquigarrow} (y, E')$ for some E'.

By assumption $\|E\|_w \succ N(y)$, and the definition 5.4.10 of $\succ$, we obtain that $\forall s \in \|E\|_w$, $\exists s' \in N(q)$, such that $\mathcal{C}^*(s, s')$ holds. Since $\mathcal{C}^*$ is a transitive closure of $\mathcal{C}$, it follows that for each state in $\|E\|_w$ there is a sequence of states in Ω, $s_1, ..., s_n$, such that

$$s_1 \in \|E\|_w, \;\; s_n \in N(y), \;\; \text{and} \;\; \mathcal{C}(s_i, s_{i+1}) \;\; \text{for} \;\; 1 \leq i < n. \tag{C.23}$$

According to the definition 5.1.2, in order to prove the lemma, we need to show that there exists a sequence of causal relationships $cr_1, ..., cr_{n-1}$, such that $cr_i : \epsilon_i$ causes ρ_i if Φ_i, and cr_i is applicable to the pair (q_i, E_i), yielding (q_{i+1}, E_{i+1}), where $q_i = p(s_i)$, $\{\rho_i\} = q_{i+1} \setminus q_i$, $E_{i+1} = (E_i \setminus \{\neg\rho_i\}) \cup \{\rho_i\}$, and $E_1 = E$.

We prove it by induction on the length of this sequence. Let $n = 2$, $q_1 = x$, and $q_2 = y$. The fact C.23 for the case $n = 2$ means that for all $s \in \|E\|_w$, there exists $s' \in N(q_2)$, such that $\mathcal{C}(s, s')$ holds. Using this and the definition of $L(q_1, q_2)$, we obtain that

$$\|E\|_w \subseteq L(q_1, q_2). \tag{C.24}$$

We need to show that there exists a causal relationship cr_1:

$$\epsilon_1 \;\; causes \;\; \rho_1 \;\; if \;\; \Phi_1,$$

applicable to (q_1, E_1) and yielding (q_2, E_2), where $\{\rho_1\} = q_2 \setminus q_1$, $E_2 = (E_1 \setminus \{\neg\rho_1\}) \cup \{\rho_1\}$. In other words, we need to show that $(x, E_1) \overset{*}{\rightsquigarrow} (y, E_2)$.

We plan to show this by proving that there exists a literal $\overset{\circ}{\epsilon_1} \in \overset{\circ}{E}$, such that $[\overset{\circ}{\epsilon_1}] \cap N(q_1) \subseteq L(q_1, q_2)$ and using the corollary 5.4.6.

The connection set $L(q_1, q_2)$ is non-empty — it contains at least the non-empty trigger set, as established by C.24. Then, the lemma 5.4.4 shows that since $L(q_1, q_2) \neq \emptyset$, there is at least one justifier literal $\overset{\circ}{\epsilon}$ such that $[\overset{\circ}{\epsilon}] \cap N(q_1) \subseteq L(q_1, q_2)$. We need to show that such a literal belongs to the justifier effect set, i.e., $\overset{\circ}{\epsilon} \in \overset{\circ}{E}$.

Let us assume the opposite and consider the case when $[\overset{\circ}{\epsilon}] \cap N(q_1) \subseteq L(q_1, q_2)$ holds only for literals $\overset{\circ}{\epsilon} \notin \overset{\circ}{E}$.

This assumption, first of all, entails that the trigger set $\|E\|_w$ is not a singleton — otherwise (if it was a singleton $\|E\|_w = \{s\}$, where $s = q_1 \cup \overset{\circ}{q_1}$), the hyper-state s would contain all the *fixed* justifier literals, and therefore, for one of them $[\overset{\circ}{\epsilon}] \cap N(q_1) \subseteq L(q_1, q_2)$ would hold.

Since the trigger set $\|E\|_w$ is not a singleton, there must be a justifier literal not in $\overset{\circ}{E}$, such that its value varies across the states in $\|E\|_w$, as required by the observation 5.4.9. This ensures that there exists a pair of states $s, s' \in \|E\|_w$ that agree with respect to all literals except the one which is not in $\overset{\circ}{E}$. In other words, $s \setminus \{\overset{\circ}{\epsilon}\} = s' \setminus \{\neg\overset{\circ}{\epsilon}\}$, where $\overset{\circ}{\epsilon} \notin \overset{\circ}{E}$.

Both states s and s' are in the trigger set, and hence, from $\{s, s'\} \subseteq \|E\|_w$ and C.24 it follows that

$$\{s, s'\} \subseteq L(q_1, q_2). \tag{C.25}$$

However, by construction of states s and s', $s' = (s \setminus \{\overset{\circ}{\epsilon}\}) \cup \{\neg\overset{\circ}{\epsilon}\}$. Therefore, $s \in [\overset{\circ}{\epsilon}] \cap N(q_1)$ and $s' \in [\neg\overset{\circ}{\epsilon}] \cap N(q_1)$. Using the lemma 5.4.7, we can see that either $s \in L(q_1, q_2)$ or $s' \in L(q_1, q_2)$ — but not both. This contradicts C.25.

It follows, as the only possibility, that there exists a literal $\overset{\circ}{\epsilon_1} \in \overset{\circ}{E}$, such that $[\overset{\circ}{\epsilon_1}] \cap N(q_1) \subseteq L(q_1, q_2)$. Then, setting $E_1 = E$ and using the corollary 5.4.6, we obtain that there is a causal relationship $cr_1 : \epsilon_1 \ causes \ \rho_1 \ if \ \Phi_1$, applicable to (q_1, E_1) and yielding (q_2, E_2), where $\{\rho_1\} = q_2 \setminus q_1$, $E_2 = (E_1 \setminus \{\neg\rho_1\}) \cup \{\rho_1\}$. Hence $(x, E_1) \overset{*}{\rightsquigarrow} (y, E_2)$.

Consider a case of length k. Let us assume that for a sequence of states in Ω, $s_1, ..., s_k$, such that $s_1 \in \|E\|_w$, $s_k \in N(q_k)$, and $\|E\|_w \succ N(q_k)$, there is a sequence of causal relationships $cr_1, ..., cr_{k-1}$, underlying $(q_1, E_1) \overset{*}{\rightsquigarrow} (q_k, E_k)$, where $q_1 = x$. Assume also that for all $s_k \in N(q_k)$ such that $\mathcal{C}^*(s, s_k)$, where $s \in \|E\|_w$, $C(s_k, s_{k+1})$ holds for some $s_{k+1} \in N(q_{k+1})$. The argument relying on the corollary 5.4.6 and observation 5.4.9 and analogous to the $n = 2$ case shows that there is a causal relationship $cr_k : \epsilon_k \ \ causes \ \ \rho_k \ \ if \ \ \Phi_k$, underlying propagation $(q_k, E_k) \rightsquigarrow (q_{k+1}, E_{k+1})$, where $E_{k+1} = (E_k \setminus \{\neg\rho_k\}) \cup \{\rho_k\}$. By transitivity of $\rightsquigarrow$, we obtain immediately $(x, E) \overset{*}{\rightsquigarrow} (q_{k+1}, E_{k+1})$.

Therefore, assuming $\|E\|_w \succ N(y)$ for some $y \in \mathcal{W}$, $\|E\|_w \subseteq N(x)$, we prove that $(x, E) \overset{*}{\rightsquigarrow} (y, E')$ for some E'.

($\Longleftarrow$) Let $\|E\|_w \subseteq N(x)$ and $(x, E) \overset{*}{\rightsquigarrow} (y, E')$, for some $x, y \in \mathcal{W}$ and E'. We need to show that $\|E\|_w \succ N(y)$.

Consider a state $s \in \|E\|_w$. Using lemma assumption, $(x, E) \overset{*}{\rightsquigarrow} (y, E')$, we can see that there exists a sequence of causal relationships

$$cr_1, ..., cr_{n-1}, cr_i : \ \epsilon_i \ causes \ \rho_i \ if \ \Phi_i,$$

underlying the propagation $(x, E) \overset{*}{\rightsquigarrow} (y, E')$. Let $\epsilon_1 \in E$. From $\cap_{s \in \|E\|_w} h(s) = \overset{\circ}{E}$ (observation 5.4.9) we obtain $h(s) \vdash \overset{\circ}{\epsilon_1}$, or $\overset{\circ}{\epsilon_1} \in s$. Then the definition 5.2.6, the definition of $\rightsquigarrow$, and simple induction capitalising on transitivity of $\mathcal{C}$ yield that $C^*(s, s')$ for some $s' \in y$. This holds for every $s \in \|E\|_w$. Therefore, using the definition of $\succ$, we obtain $\|E\|_w \succ N(y)$ as desired.

Together, the results established in ($\Longrightarrow$) and ($\Longleftarrow$) prove the lemma.

■

Theorem 5.4.14 $Res_{RD\mathcal{L}}(w, a) = Res_{\Omega}(w, a)$.

Proof:

($\Longleftarrow$) Consider a state $y \in Res_{\Omega}(a, \ w)$. We need to show that $y \in Res_{RD\mathcal{L}}(a, \ w)$.

The case of a zero-length causal propagation is trivial: if there are no $\mathcal{C}$-links from a state $y \in Res_{\Omega}(a, \ w)$, then, using the definition of 5.2.6, we observe that there are no causal relationships applicable to (y, E). Moreover, by definition 5.4.12, the state y satisfies all domain constraints in D. Therefore, $y \in Res_{RD\mathcal{L}}(a, \ w)$ (in fact, both resultant sets are singletons because E is just a set of literals).

The case of a k-length causal propagation ($k > 0$) follows immediately from the lemma 5.4.13. If $y \in Res_{\Omega}(a, \ w)$ then the lemma ensures that $(x, E) \overset{*}{\rightsquigarrow} (y, E')$ where $\|E\|_w \subseteq N(x)$ for some E'. The fact $\|E\|_w \subseteq N(x)$ yields $x = (w \setminus C) \cup E$ (by properties of the PMA ordering). In addition, if a state $y \in Res_{\Omega}(a, \ w)$, the set $T^*(\|E\|_w, y)$ is final and, hence, there are no $\mathcal{C}$-links from some elements of $T^*(\|E\|_w, y)$. Then, using the definition of 5.2.6 and corollary 5.4.6, we obtain that there are no causal relationships

applicable to (y, E') for the E' obtained by propagation $(x, E) \overset{*}{\rightsquigarrow} (y, E')$. Again, the definition 5.4.12 ensures that the state y satisfies all domain constraints in D. Using the definition 5.1.3, we obtain that $y \in Res_{RD\mathcal{L}}(a,\ w)$, and therefore,

$$Res_{\Omega}(a,\ w) \subseteq Res_{RD\mathcal{L}}(a,\ w).$$

($\Longrightarrow$) Consider a state $y \in Res_{RD\mathcal{L}}(a,\ w)$. We need to show that $y \in Res_{\Omega}(a,\ w)$.

Again, the case of a zero-length causal propagation is trivial: if there are no causal relationships applicable to (y, E), where $y \in Res_{RD\mathcal{L}}(a,\ w)$, then there are no $\mathcal{C}$-links from y (follows from the definition 5.2.6). Hence, the set $T^*(\|E\|_w, y) = \|E\|_w$ is final. Therefore, by definition 5.4.12, $y \in Res_{\Omega}(a,\ w)$.

The case of a k-length causal propagation ($k > 0$) follows immediately from the lemma 5.4.13. If $y \in Res_{RD\mathcal{L}}(a,\ w)$, then the lemma ensures that $\|E\|_w \succ N(y)$. In addition, if there are no causal relationships applicable to (y, E') for the E' obtained by propagation $(x, E) \overset{*}{\rightsquigarrow} (y, E')$, then there are no $\mathcal{C}$-links from some elements of the set $T^*(\|E\|_w, y)$ (definition 5.2.6 and corollary 5.4.6), and it is final. Therefore, by definition 5.4.12, $y \in Res_{\Omega}(a,\ w)$, and

$$Res_{RD\mathcal{L}}(a,\ w) \subseteq Res_{\Omega}(a,\ w).$$

Together, the results established in ($\Longrightarrow$) and ($\Longleftarrow$) prove that

$$Res_{RD\mathcal{L}}(a,\ w) = Res_{\Omega}(a,\ w).$$

■

Corollary 5.4.15 *For any states $w \in \mathcal{W}$ and $q \in \mathcal{W}$, $\|E\|_w \succ N(q)$ if and only if $\gamma(\|E\|_w) \overset{*}{\rightharpoonup} \gamma(T^*(\|E\|_w, q))$.*

Proof:

($\Longrightarrow$) Let $\|E\|_w \succ N(q)$ for some w and q.

The corollary assumption $\|E\|_w \succ N(q)$ means that there is a sequence of causally triggered hyper-neighbourhoods $N(q_1), \ldots, N(q_n) = N(q)$ for n states $q_1, \ldots, q_n = q$, where $n \geq 1$, such that $\|E\|_w \succ N(q_i)$, for all i, where $1 \leq i \leq n$.

We intend to show, by induction on n, that $\gamma(\|E\|_w) \overset{*}{\rightharpoonup} \gamma(T^*(\|E\|_w, q_i))$ for all i, where $1 \leq i \leq n$.

Let $n = 1$. By definition 5.4.10 of $\succ$, for all $s \in \|E\|_w$ there exists $s' \in N(q_1)$, such that $\mathcal{C}(s, s')$. Let us consider the maximal subset of $N(q_1)$, denoted z, containing all states s' such that $\mathcal{C}(s, s')$. In other words, for all $s' \in z$, $\mathcal{C}(s, s')$ holds, and there is no state $s'' \in N(q_1) \setminus z$, such that $\mathcal{C}(s, s')$, where $s \in \|E\|_w$.

By definition 5.4.2 of the traced set, for all elements $s' \in T(\|E\|_w, q_1)$ there exists an element of the trigger set $s \in \|E\|_w$, such that $\mathcal{C}(s, s')$. Therefore, the constructed set z is $T(\|E\|_w, q_1)$.

Hence, there are no hyper-states in $\|E\|_w$ without an out-going $\mathcal{C}$-link to some hyper-state in $T(\|E\|_w, q_1)$, and there are no hyper-states in $T(\|E\|_w, q_1)$ without an incoming $\mathcal{C}$-link from some hyper-state in $\|E\|_w$.

Then, by Definition 5.3.2 of $\rightharpoonup$, we obtain

$$\gamma(\|E\|_w) \rightharpoonup \gamma(T(\|E\|_w, q_1)) \quad \text{or} \quad \gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(T(\|E\|_w, q_1)).$$

Let $i > 1$, and let us assume that

$$\gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(T^*(\|E\|_w, q_i)). \tag{C.26}$$

We need to show that $\gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(T^*(\|E\|_w, q_{i+1}))$.

By using the argument employed for the case $n = 1$, where the subset $T^*(\|E\|_w, q_i)$ of the hyper-neighbourhood $N(q_i)$ plays the role of the trigger set $\|E\|_w$, we obtain that

$$\gamma(T^*(\|E\|_w, q_i)) \rightharpoonup \gamma(T^*(\|E\|_w, q_{i+1})). \tag{C.27}$$

Using C.26, C.27, and transitivity of the relation $\stackrel{*}{\rightharpoonup}$, we conclude that

$$\gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(T^*(\|E\|_w, q_{i+1})).$$

Therefore, $\gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(T^*(\|E\|_w, q))$, where $q = q_n$.

($\Longleftarrow$) Let $\gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(T^*(\|E\|_w, q))$ for some w and q.

This assumption and an induction similar to the one used in ($\Longrightarrow$), entail that for all $s \in \|E\|_w$, $\mathcal{C}^*(s, s')$ holds for some state $s' \in T^*(\|E\|_w, q))$.

Then, by the definition 5.4.10 of $\succ$, we obtain that $\|E\|_w \succ N(q)$.

Together, the results established in ($\Longrightarrow$) and ($\Longleftarrow$) prove the corollary.

■

Theorem 5.4.17 $Res_{\Omega}(w, a) = Res_{\Gamma}(w, a)$.

Proof:

($\Longrightarrow$) Consider a state $y \in Res_{\Omega}(a,\ w)$. We need to show that $y \in Res_{\Gamma}(a,\ w)$.

If $y \in Res_{\Omega}(a,\ w)$, then by definition 5.4.12 of a successor state with respect to the hyper-state space semantics, $\|E\|_w \succ N(y)$ and the set $T^*(\|E\|_w, y)$ is final.

By corollary 5.4.15, $\gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(T^*(\|E\|_w, y))$.

In other words, $\gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(z)$ for some $z = T^*(\|E\|_w, y)$, where $z \subseteq N(y)$, and $\gamma(z)$ is final.

Therefore, by definition 5.4.16, the state y is a successor state with respect to the power-state space semantics, $y \in Res_{\Gamma}(a,\ w)$, and

$$Res_{\Omega}(w, a) \subseteq Res_{\Gamma}(w, a).$$

($\Longleftarrow$) Consider a state $y \in Res_{\Gamma}(a,\ w)$. We need to show that $y \in Res_{\Omega}(a,\ w)$.

If $y \in Res_{\Gamma}(a,\ w)$, then by definition 5.4.16 of a successor state with respect to the power-state space semantics, $\gamma(\|E\|_w) \stackrel{*}{\rightharpoonup} \gamma(z)$ for some set $z \subseteq N(y)$, and $\gamma(z)$ is final.

For any element $s' \in z$, $\mathcal{C}^*(s, s')$ holds for some $s \in \|E\|_w$ — in other words, there is an incoming (transitive) $\mathcal{C}^*$-link from a state in the trigger set to each element s' of z. This observation and the fact that $z \subseteq N(y)$ entail that $\|E\|_w \succ N(y)$, by the definition 5.4.10 of $\succ$.

Moreover, each element $s' \in z$ is, by the definition 5.4.3 of transitively traced sets, an element of $T^*(\|E\|_w, y)$ as well. It follows that $z \subseteq T^*(\|E\|_w, y)$. Since the set z is final and is a subset of the set $T^*(\|E\|_w, y)$, the latter is final too (it contains at least on final hyper-state).

Therefore, by definition 5.4.12, the state y is a successor state with respect to the hyper-state space semantics, $y \in Res_{\Omega}(a,\ w)$, and

$$Res_{\Gamma}(w, a) \subseteq Res_{\Omega}(w, a).$$

Together, the results established in ($\Longrightarrow$) and ($\Longleftarrow$) prove that

$$Res_{\Omega}(a,\ w) = Res_{\Gamma}(a,\ w).$$

■

Appendix D

Proofs for Chapter 6

Lemma 6.3.1 *If the relation G satisfies the conditions $(G_1) - (G_3)$, then for each $w \in \mathcal{W}$, the ordering $<_{w,G}$ satisfies conditions $(M_1) - (M_3)$.*

Proof:

Let the relation G satisfy conditions $(G_1) - (G_3)$.

a) We show first that the preference relation $<_{w,G}$ defined by definition 6.2.5 satisfies condition (M_1).

Consider states $p, q, x \in \mathcal{W}$ such that $p <_{w,G} q$ and $q <_{w,G} x$. By definition 6.2.5, the fact $p <_{w,G} q$ yields $\forall e \in \mathcal{E}$, such that $p, q \in [e]$, $\neg G(e, w, q)$ and $\exists a \in \mathcal{E}$, such that $p, q \in [a], G(a, w, p)$. Analogously, the same definition and the fact $q <_{w,G} x$ yield $\forall e' \in \mathcal{E}$, such that $q, x \in [e']$, $\neg G(e', w, x)$ and $\exists a' \in \mathcal{E}$, such that $q, x \in [a'], G(a', w, q)$.

Consider now an action $e'' \in \mathcal{E}$, such that $p, x \in e''$. Since, by lemma assumption, the relation G satisfies condition (G_1), we can apply its syntactically expanded variant:

$$(\exists a \in \mathcal{E}, p, q \in [a], G(a, w, p)) \wedge (\forall e \in \mathcal{E}, p, q \in [e], \neg G(e, w, q))$$

$$\wedge (\exists a' \in \mathcal{E}, q, x \in [a'], G(a', w, q)) \wedge (\forall e' \in \mathcal{E}, q, x \in [e'], \neg G(e', w, x))$$

imply

$$(\exists a'' \in \mathcal{E}, p, x \in [a''], G(a'', w, p)) \wedge (\forall e'' \in \mathcal{E}, p, x \in [e''], \neg G(e'', w, x)).$$

The implication and definition 6.2.5 yields $p <_{w,G} x$. Hence, the condition M_1 is satisfied.

b) Now we intend to show that the preference relation $<_{w,G}$ satisfies condition (M_2).

Consider an arbitrary action e and states $p, q \in [e]$, such that

$$\neg\exists x \in [e], x \neq p,\ x <_{w,G} p, \tag{D.1}$$

and assume that

$$\forall e \in \mathcal{E}, p, q \in [e], \exists y \in [e], y <_{w,G} q. \tag{D.2}$$

These two assumptions set the premises of the condition (M_2). By definition 6.2.5, the fact D.2 yields that state q is never selected by $\{p, q\}$-covering actions:

$$\forall e \in \mathcal{E},\ \text{such that } p, q \in [e], \neg G(e, w, q). \tag{D.3}$$

Now we only need to show that $\exists a' \in \mathcal{E}$, such that $p, q \in [a'], G(a', w, p)$ (in other words, state p is selected at least once by a $\{p, q\}$-covering action). To do so we assume the opposite:

$$\forall a' \in \mathcal{E} \text{ such that } p, q \in [a'], \neg G(a', w, p). \tag{D.4}$$

This will be true for all such actions, including e, therefore, $\neg G(e, w, p)$. Since action e, invoked at w, results neither in p nor in q, the condition (G_3) requires that $G(e, w, h)$ for some other state $h \in [e]$. Then the condition (G_2') ensures that invocation of any action e'' such that both states $h, p \in e''$, cannot result in p, that is, $\neg G(e'', w, p)$ (it follows from $G(e, w, h) \wedge \neg G(e, w, p)$ and $h, p \in [e] \cap [e'']$). Thus, we established that

$$\forall e'' \in \mathcal{E}, h, p \in [e''], \neg G(e'', w, p).$$

The latter and the fact $G(e, w, h)$, by definition 6.2.5, will yield $h <_{w,G} p$. This conclusion contradicts the earlier assumption D.1 — note that $h \in [e]$. Therefore, the assumption D.4 is wrong and there exists an action $a' \in \mathcal{E}$, such that $p, q \in [a'], G(a', w, p)$.

The last conclusion, alongside with D.3, leads to $p <_{w,G} q$, following definition 6.2.5. Hence, (M_2).

c) Finally, we show that the relation $<_{w,G}$ satisfies condition (M_3).

Let $p <_{w,G} q$. Then, by definition 6.2.5,

$$\forall e \in \mathcal{E} \text{ such that } p, q \in [e],\ \ \neg G(e, w, q),$$

and

$$\exists e \in \mathcal{E} \text{ such that } p, q \in [e], \;\; G(e, w, p). \tag{D.5}$$

The latter fact alone ensures that $\neg\exists x \in [e], x <_w p$. To verify this, assume the opposite $\exists x \in [e], x <_w p$. This would require, by definition 6.2.5, $\forall e \in \mathcal{E}, x, p \in [e], \neg G(e, w, p)$, contradicting D.5. Therefore, $\exists e \in \mathcal{E}, p, q \in [e]$, such that $\neg\exists x \in [e], x <_w p$, establishing (M_3).

■

Lemma 6.3.2 *If each ordering $<_w$ for $w \in \mathcal{W}$ satisfies conditions $(M_1) - (M_3)$, then the relation $G_<$ satisfies the conditions $(G_1) - (G_3)$.*

Proof:

Let the relation $<_w$ satisfy condition $(M_1) - (M_3)$.

a) We show first that the invocation relation $G_<$ defined in 6.2.4 satisfies condition (G_1).

Consider states $p, q, x \in \mathcal{W}$ such that on one hand,

$$\forall e \in \mathcal{E}, p, q \in [e], \;\; \neg G_<(e, w, q) \tag{D.6}$$

and

$$\exists a \in \mathcal{E}, p, q \in [a], \;\; G_<(a, w, p), \tag{D.7}$$

and on the other hand,

$$\forall e' \in \mathcal{E}, q, x \in [e'], \;\; \neg G_<(e', w, x) \tag{D.8}$$

and

$$\exists a' \in \mathcal{E}, p, q \in [a'], \;\; G_<(a', w, q). \tag{D.9}$$

In other words, we assumed that the left-hand side of the implication in the condition (G_1) holds.

By definition 6.2.4, the fact D.7 means that

$$\neg\exists x \in [a], p, q \in [a] \text{ such that } x <_w p. \tag{D.10}$$

The fact D.6 yields

$$\forall e \in \mathcal{E}, p, q \in [e], \exists y \in [e], y <_w q. \tag{D.11}$$

Then the condition (M_2) ensures that the facts D.10 and D.11 imply $p <_w q$.

Analogously, the facts D.8 and D.9 and condition (M_2) lead to $q <_w x$. Now, the condition (M_1) and facts $p <_w q$ and $q <_w x$ yield $p <_w x$. This conclusion and the definition 6.2.4 guarantee that

$$\forall e \in \mathcal{E}, p, x \in [e], \neg G_<(e, w, x),$$

establishing one conjunct on the right-hand side of condition (G_1).

The condition (M_3) requires that, since $p <_w x$, then $\exists e' \in \mathcal{E}, p, x \in [e']$, such that $\neg\exists z \in [e'], z \neq p, z <_w p$ — ensuring that p is a minimal state in some $[e']$. Hence,

$$\exists e' \in \mathcal{E}, p, x \in [e'], \text{ such that } G_<(e', w, p),$$

proving the other conjunct on the right-hand side of condition (G_1). Hence, (G_1).

b) Now let us show that the relation $G_<$ satisfies condition (G_2), in other words, that for all $\{p, q\}$-covering actions $e', e'' \in \mathcal{E}$ and states $w \in \mathcal{W}$, the fact $G_<(e', w, p) \wedge G_<(e'', w, q)$ implies $G_<(e', w, q)$.

Assume

$$G_<(e', w, p) \wedge G_<(e'', w, q), \text{ where } p, q \in [e'] \cap [e''] \tag{D.12}$$

and the opposite of the desired implication, $\neg G_<(e', w, q)$. By definition 6.2.4, the latter implies that some other state in $[e']$ is preferred to q:

$$\forall e' \in \mathcal{E}, p, q \in [e'], \exists y \in [e'], \text{ such that } y <_w q. \tag{D.13}$$

From the fact $G_<(e', w, p)$ — the first conjunct in D.12 — we obtain $\neg\exists x \in [e']$ such that $x \neq p$ and $x <_w p$. This and fact D.13 yield, using condition (M_2), $p <_w q$. In other words, the minimal (in $[e']$) state p is preferred to the non-minimal (in $[e']$) state q.

The obtained fact $p <_w q$ would imply, by definition 6.2.4, that invocation of any action e'' such that $p, q \in [e'']$, cannot result in state q, meaning $\neg G_<(e'', w, q)$. This contradicts the assumption $G_<(e'', w, q)$ — the second conjunct in D.12.

Therefore, the assumption $\neg G_<(e', w, q)$ was wrong, and

$$G_<(e', w, p) \wedge G_<(e'', w, q) \supset G_<(e', w, q)$$

for all $p, q \in [e'] \cap [e'']$. Hence, (G_2).

c) What remains to be shown is that $G_<$ satisfies condition (G_3), in other words, that for all $e \in \mathcal{E}, w \in \mathcal{W}$, there exists $p \in [e]$, such that $G_<(e, w, p)$.

Assume the opposite: $\exists e \in \mathcal{E}, w \in \mathcal{W}$, such that $\forall p \in [e]$, the fact $\neg G_<(e, w, p)$ holds. Consider any state $p_1 \in [e]$. By definition 6.2.4, the fact $\neg G_<(e, w, p_1)$ would mean $\exists p_2 \in [e], p_2 <_w p_1$. Since $p_2 \in [e]$, the fact $\neg G_<(e, w, p_2)$ holds as well. Therefore, continuing this inductive process, we obtain $\exists p_n \in [e], p_n <_w p_{n-1}$. Since the set $[e]$ is finite, such a process will require that $\exists p_i \in [e], p_i <_w p_n$, where $1 \leq i < n$. However, the transitivity condition (M_1) ensures that $\forall i, 1 \leq i < n, p_n <_w p_i$. The contradiction shows that $\forall e \in \mathcal{E}, w \in \mathcal{W}, \exists p \in [e]$, such that $G_<(e, w, p)$. Hence, (G_3).

■

Lemma 6.3.3 $G_{<_{w,G}}(e, w, r)$ *if and only if* $G(e, w, r)$.

Proof:

($\Longleftarrow$) First, we establish that

$$G(e, w, r) \subseteq G_{<_{w,G}}(e, w, r).$$

Let $G(e, w, r)$. We need to show $G_{<_{w,G}}(e, w, r)$. Assume the opposite: $\neg G_{<_{w,G}}(e, w, r)$. This means that, according to the definition 6.2.4, there exists $x \in [e]$, such that $x \neq r$ and $x <_{w,G} r$. Then by definition 6.2.5, for all $\{x, r\}$-covering actions $e' \in \mathcal{E}$ (in other words, $x, r \in [e']$), $\neg G(e', w, r)$ holds and there exists an $\{x, r\}$-covering action $a \in \mathcal{E}$ (in other words, $x, r \in [a]$), such that $G(a, w, x)$. The former fact alone yields $\neg G(e, w, r)$ (noting that $r, x \in [e]$). This conclusion contradicts the lemma assumption $G(e, w, r)$. Therefore, $G_{<_{w,G}}(e, w, r)$ holds, and

$$G(e, w, r) \subseteq G_{<_{w,G}}(e, w, r). \tag{D.14}$$

($\Longrightarrow$) Now let us show $G_{<_{w,G}}(e, w, r) \subseteq G(e, w, r)$.

Let $G_{<_{w,G}}(e, w, r)$. By definition 6.2.4,

$$\neg \exists x \in [e], \text{ such that } x \neq r \text{ and } x <_{w,G} r. \tag{D.15}$$

Or, alternatively, for all $x \in [e]$, $\neg(x <_{w,G} r)$. Applying definition 6.2.5 to $\neg(x <_{w,G} r)$, we obtain that for all $x \in [e]$,

$$\neg((\forall e' \in \mathcal{E}, x, r \in [e'], \neg G(e', w, r)) \wedge (\exists a \in \mathcal{E}, x, r \in [a], G(a, w, x))).$$

It follows that for all $x \in [e]$,

$$\neg(\forall e' \in \mathcal{E}, x, r \in [e'], \neg G(e', w, r)) \vee \neg(\exists a \in \mathcal{E}, x, r \in [a], G(a, w, x)),$$

which is equivalent to the following: for all $x \in [e]$,

$$(\exists e' \in \mathcal{E}, x, r \in [e'], G(e', w, r)) \vee (\forall a \in \mathcal{E}, x, r \in [a], \neg G(a, w, x)). \tag{D.16}$$

We need to prove that $G(e, w, r)$. Let us assume the opposite $\neg G(e, w, r)$. The question we pose now is whether $G(e, w, x)$ or not. We intend to show that either outcome is not possible, leading to a contradiction.

If $G(e, w, x)$, then by the condition (G_2'), for all $e'', r, x \in [e'']$,

$$\neg G(e, w, r) \wedge G(e, w, x) \supset \neg G(e'', w, r).$$

This implication would eliminate a possibility that the first disjunct in D.16 is true. At the same time, $G(e, w, x)$ means that the second disjunct in D.16 is false as well, resulting in a contradiction.

Since the assumption $G(e, w, x)$ is incorrect, the only possibility remaining is that $\neg G(e, w, x)$. Since invocation of action e at state w results in neither r nor x, the condition (G_3) requires that there exists $h \in [e], G(e, w, h)$. Again, the condition (G_2') is applied: for all $e'', r, h \in [e'']$,

$$\neg G(e, w, r) \wedge G(e, w, h) \supset \neg G(e'', w, r).$$

Now, by the definition 6.2.5, we can obtain $h <_{w,G} r$. Both states $h, r \in [e]$, and therefore, the obtained fact contradicts D.15.

Thus, neither $G(e, w, x)$ nor $\neg G(e, w, x)$. This contradiction implies $G(e, w, r)$, and hence,

$$G_{<_{w,G}}(e, w, r) \subseteq G(e, w, r). \tag{D.17}$$

The established containments D.14 and D.17 yield the desired identity

$$G(<_{w,G}, e, w, r) = G(e, w, r).$$

■

Lemma 6.3.4 *For each ordering* $<_w$, $p <_{w,G_<} q$ *if and only if* $p <_w q$.

Proof:

($\Longrightarrow$) First, we show that for any two states $p, q \in \mathcal{W}$, if $p <_{w,G_<} q$ then $p <_w q$.

Let $p <_{w,G_<} q$. Then by definition 6.2.5 of new ordering $<_{w,G_<}$,

$$\forall e \in \mathcal{E}, \text{ such that } p, q \in [e], \ \neg G_<(e, w, q) \tag{D.18}$$

and

$$\exists a \in \mathcal{E}, \text{ such that } p, q \in [a], \ G_<(a, w, p).$$

The latter fact and the definition 6.2.4 of new relation $G_<$ imply

$$\neg\exists x \in [a], \text{ such that } x \neq p \text{ and } x <_w p. \tag{D.19}$$

This means that state p is a minimal element of the set $[a]$. We intend to show now that the state q is not a minimal element of set $[a]$.

The fact D.18 and the definition 6.2.4 imply

$$\forall e \in \mathcal{E}, p, q \in [e], \exists y \in [e], y <_w q, \tag{D.20}$$

establishing that the state q is not a minimal element of any set $[e]$, where e is a $\{p, q\}$-covering action, for example the action a.

Hence, the facts D.19, D.20 and the condition (M_2) establish that $p <_w q$. Therefore, if $p <_{w,G_<} q$ then $p <_w q$.

($\Longrightarrow$) We need to show that for any two states $p, q \in \mathcal{W}$, if $p <_w q$ then $p <_{w,G_<} q$.

Let $p <_w q$, and let us assume the opposite of $p <_{w,G_<} q$. In other words, $\neg(p <_{w,G_<} q)$. Applying definition 6.2.5 to $\neg(p <_{w,G_<} q)$, we obtain

$$\neg((\forall e \in \mathcal{E}, p, q \in [e], \neg G_<(e, w, q)) \wedge (\exists a \in \mathcal{E}, p, q \in [a], G_<(a, w, p))).$$

It follows that

$$\neg(\forall e \in \mathcal{E}, p, q \in [e], \neg G_<(e, w, q)) \vee \neg(\exists a \in \mathcal{E}, p, q \in [a], G_<(a, w, p)),$$

which is equivalent to

$$(\exists e \in \mathcal{E}, p, q \in [e], G_<(e, w, q)) \vee (\forall a \in \mathcal{E}, p, q \in [a], \neg G_<(a, w, p)) \qquad \text{(D.21)}$$

Consider the first disjunct in D.21. It states that there exists a $\{p, q\}$-covering action e such that the state q is selected by the relation $G_<$, namely $G_<(e, w, q)$. This means, by definition 6.2.4, that

$$\neg\exists x \in [e], \text{ such that } x \neq q \text{ and } x <_w \ q.$$

This, however, contradicts the lemma assumption $p <_w q$. Hence, the first disjunct in D.21 cannot be true.

Consider the second disjunct in D.21. It states that for all $\{p, q\}$-covering actions e, the state p is never selected by the relation $G_<$, namely $\neg G_<(e, w, p)$. However, the fact $p <_w q$ and the condition (M_3) require that the state p is $<_w$-minimal in $[e']$ for some $\{p, q\}$-covering action e', or in other words, $G_<(e', w, p)$. This contradicts $\neg G_<(e, w, p)$ for all $\{p, q\}$-covering actions e, making the second disjunct in D.21 false as well. The obtained contradiction means that $p <_{w,G_<} q$.

Therefore, $p <_{w,G_<} q$ if and only if $p <_w q$.

■

Lemma 6.4.4 *A strongly respectful system is trivial.*

Proof:

Consider a strongly respectful action system. Assume it is not trivial. Then there exists at least one transition chain. Let $p_1, p_2, ..., p_n, (q)$, $n \geq 1$, be a transition chain for some state w: $G(e, w, p_1)$. In a respectful action system, the last element of any finite chain must be a member of $\mathcal{D}$. Since $q \in \mathcal{D}$ then it has to be respected by any pair in any transition chain: for example, in the transition chain $p_1, p_2, ..., p_n, (q)$. The discreteness property $(\mathcal{O}_3)$ of the ordering $<_q$ requires, as mentioned in the Chapter 2, that any state q is the single minimal element with respect to an ordering $<_q$ centered on itself. Clearly,

for any p_i, $q <_q p_i$ and $\neg(p_i <_q q)$. Hence, by lemma 6.4.1, q is not respected by any pair p_i, q, for all $i, 1 \leq i \leq n$.

Therefore, by contradiction, a strongly respectful system is trivial.

■

Lemma 6.4.5 *In a weakly respectful system, for any pair p, q such that $C^*(p, q)$, and state q is stable, and for every action e, there is no $G(e, q, p)$.*

Proof:

Consider a pair p, q such that $C^*(p, q)$, where q is stable. Since the system is respectful, $q \in \mathcal{D}$. Then q has to be respected by any pair in a transition chain starting with s, where $G(e, q, s)$. Clearly, q is not respected by the pair p, q (the discreteness property ($\mathcal{O}_3$) of the ordering $<_q$). Hence, $s \neq p$. In other words, there is no $G(e, q, p)$.

■

Lemma 6.4.6 *In a weakly respectful system, any two states w and s that share a causal link (p_1, p_2), agree on all state variables f that change values between p_1 and p_2: $p_1(f) \neq p_2(f)$ implies $w(f) = s(f)$.*

Proof:

Consider a pair p_1, p_2 such that $w_1, ..., p_1, p_2, ..., (q)$ is a transition chain for the state w, invoked by $G(e, w, w_1)$, and $s_1, ..., p_1, p_2, ..., (q)$ is a transition chain for the state s, invoked by $G(e', s, s_1)$. States w_1 and s_1 do not have to be different. Since the action system is respectful, the pair p_1, p_2 respects both w and s. By definition,

$$p_1(f) \neq p_2(f) \supset p_1(f) = w(f)$$

$$p_1(f) \neq p_2(f) \supset p_1(f) = s(f)$$

By lemma assumption, $p_1(f) \neq p_2(f)$. It follows immediately that $w(f) = s(f)$.

■

Lemma 6.4.7 *In a weakly respectful system where every state has n state variables, the number of states allowed to share causal chains of length k is restricted from above by $2^{(n-k)} - 1$.*

Proof:

In a respectful system, a causal chain of length k has to use up all k state variables. Then $(n - k)$ is the number of state variables allowed to vary. All possible combina-

tions amount to $2^{(n-k)}$ states. One of these states is the beginning of the chain. So the maximum number of sharing states is $2^{(n-k)} - 1$.

■

Appendix E

Proofs for Chapter 7

Lemma 7.2.2 *For a state $w \in \mathcal{W}$ and an action law $\langle C, a, E\rangle$,*

$$\gamma(\|E\|_w) \in min(\ll_{\gamma(N(w))}, [E]^{\Gamma}).$$

Proof:

Let state q be the nearest state to the initial state w among the post-condition states $[E]$, in terms of the PMA ordering $\prec_w$. In other words, $q \in min(\prec_w, [E])$.

Let us recall now that (by definition) the trigger set $\|E\|_w$ is always contained in the hyper-neighbourhood $N(q)$ of the state q, that is $\|E\|_w \subset N(q)$.

Consider the information (power-) state $\gamma(\|E\|_w)$ that corresponds to the partial hyper-state $\gamma_q(\|E\|_w)$. Its projection is precisely q which is a $\prec_w$-minimal state in $[E]$. In other words, there is no state $p \in [E]$ distinct from q such that $p \prec_w q$. Consider an information state $\gamma(z)$ in $[E]^{\Gamma}$ that corresponds to some set z in the hyper-neighbourhood $N(p)$ distinct from $N(q)$. We know that $p \prec_w q$ does not hold for any state $p \in [E]$. Then, by construction of our projection function, it follows that $\gamma(z) \ll_{\gamma(N(w))} \gamma_q(\|E\|_w)$ also does not hold for any information state $\gamma(z)$ that corresponds to the partial hyper-state $\gamma_p(z)$, where $p \neq q$.

Hence, the state $\gamma(\|E\|_w)$ would be preferred in terms of $\ll_{\gamma(N(w))}$ to any other information state $\gamma(z)$ that corresponds to the partial hyper-state $\gamma_p(z)$, where $p \neq q$.

This leaves only those information space contenders for minimality that correspond to the same hyper-neighbourhood $N(q)$. Among them, however, the state $\gamma(\|E\|_w)$

would be minimal because the divergent change between w and the partial hyper-state $\gamma_q(\|E\|_w)$ is empty by definition. To verify this, we note that $Obs(q, w) = \overset{\circ}{E}$ and $Just(\gamma_q(\|E\|_w), w) = \overset{\circ}{E}$. On the other hand, all other divergent change sets corresponding to partial hyper-states with at least one different justifier literal, are non-empty.

■

Bibliography

[1] John Anderson. The Problem of Causation. *Australasian Journal of Psychology and Philosophy* **16**: 127 – 142, 1938.

[2] Andrew B. Baker. Nonmonotonic Reasoning in the Framework of Situation Calculus. *Artificial Intelligence* **49**: 5 – 23, 1991.

[3] Chitta Baral. Reasoning about actions: Non-deterministic effects, Constraints, and Qualification. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, 2017 – 2026, 1995.

[4] Gerhard Brewka and Joachim Hertzberg. How to Do Things with Worlds: on Formalizing Actions and Plans. *Journal of Logic and Computation* Volume 3, **5**: 517 – 532, 1993.

[5] Mario Bunge. *Causality. The Place of the Causal Principle in Modern Science*. Harvard University Press, Cambridge, Massachusetts, 1959.

[6] Mario Bunge. *Treatise on Basic Philosophy. Ontology I: The Furniture of the World*. D. Reidel Publishing Company, Dordrecht-Holland/Boston-U.S.A., 1977.

[7] Donald Davidson. Causal Relations. In Ernest Sosa and Michael Tooley (editors), *Causation*, Oxford University Press, 1993.

[8] Curt John Ducasse. On the Nature and the Observability of the Causal Relation. In Ernest Sosa and Michael Tooley (editors), *Causation*, Oxford University Press, 1993.

[9] Umberto Eco. *The Name of the Rose*. Minerva, 1992.

[10] Umberto Eco. Signs of the Times. In Catherine David, Frederic Lenoir, Jean-Philippe de Tonnac (editors), *Conversations about the End of Time*, Penguin Books, 2000.

[11] Richard Fikes and Nils J. Nilsson. STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving. *Artificial Intelligence* **2**: 189 – 208, 1971.

[12] Antony Galton. Time and Change for AI. In Dov M. Gabbay, C. J. Hogger, and J. A. Robinson (editors), *Handbook of Logic in Artificial Intelligence and Logic Programming*, Volume 4, Epistemic and Temporal Reasoning, Clarendon Press, Oxford, 1995.

[13] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Bradford Books, MIT Press, Cambridge Massachusetts, 1988.

[14] Hector Geffner. Causality, Constraints and the Indirect Effects of Actions. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, Nagoya, 555 – 560, 1997.

[15] Michael Gelfond and Vladimir Lifschitz. Representing Action and Change by Logic Programs. *The Journal of Logic Programming* **17**: 301 – 322, 1993.

[16] Michael R. Genesereth and Nils J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, 1987.

[17] Matthew L. Ginsberg and David E. Smith. Reasoning about Action I: A Possible Worlds Approach. *Artificial Intelligence* **35**:165 – 195, 1988.

[18] Enrico Giunchiglia, G. Neelakantan Kartha and Vladimir Lifschitz. Actions with Indirect Effects (Extended Abstract). In *Working Notes of the AAAI Spring Symposium on Extending Theories of Action*, 80 – 85, 1995.

[19] Enrico Giunchiglia and Vladimir Lifschitz. Dependent Fluents. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, 1964 – 1969, 1995.

[20] Joakim Gustafsson and Patrick Doherty. Embracing Occlusion in Specifying the Indirect Effects of Actions. In L. Aiello, J. Doyle, and S. Shapiro (editors), *Proceedings of the Fifth International Conference on Knowledge Representation and Reasoning*, Morgan-Kaufmann, 87 – 98, 1996.

[21] G. Neelakantan Kartha. Soundness and Completeness Theorems for Three Formalizations of Action. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 724 – 729, 1993.

[22] G. Neelakantan Kartha. On the Range of Applicability of Baker's Approach to the Frame Problem. In *Proceedings of the Third Symposium on Logical Formalizations of Commonsense Reasoning*, 1996.

[23] G. Neelakantan Kartha and Vladimir Lifschitz. Actions with Indirect Effects (Preliminary Report). In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning*, Bonn, 341 – 350, 1994.

[24] Sarit Kraus, Daniel Lehmann and Menachem Magidor. Nonmonotonic Reasoning, Preferential Models and Cumulative Logics, *Artificial intelligence*, **44**: 167 – 207, 1991.

[25] David Lewis. Causation. *Journal of Philosophy*, **70**: 556 – 567, 1973.

[26] David Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.

[27] Vladimir Lifschitz. Computing Circumscription. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 121 – 127. Los Angeles, 1985.

[28] Vladimir Lifschitz. On the Semantics of STRIPS. In *Proceedings of the Workshop on Planning and Reasoning about Action*, Timberline, OR, 1986.

[29] Vladimir Lifschitz. Frames in the Space of Situations. *Artificial Intelligence* **46**: 365 – 376, 1990.

[30] Vladimir Lifschitz. Nested Abnormality Theories. *Artificial Intelligence* **74**: 351 – 365, 1995.

[31] Vladimir Lifschitz. Two Components of an Action Language. In *Proceedings of the Third Symposium on Logical Formalizations of Commonsense Reasoning*, Stanford, 1996.

[32] Fangzhen Lin and Raymond Reiter. State Constraints Revisited. In *Journal of Logic and Computation*, 4(5): 655 – 678, 1994.

[33] Fangzhen Lin. Embracing Causality in Specifying the Indirect Effects of Actions. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, 1985 – 1991, 1995.

[34] Charles Lineweaver. The Origin of the Universe. In *Newton. Graphic Science*, September - October 2000, 34 – 71, 2000.

[35] John L. Mackie. *The Cement of the Universe*. Oxford, 1974.

[36] John L. Mackie. Causes and Conditions. In Ernest Sosa and Michael Tooley (editors), *Causation*, Oxford University Press, 1993.

[37] Norman McCain and Hudson Turner. A Causal Theory of Ramifications and Qualifications. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, 1978 – 1984, 1995.

[38] Norman McCain and Hudson Turner. Causal Theories of Action and Change. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Innovative Applications of Artificial Intelligence Conference*, Providence, Rhode Island, AAAI Press, The MIT Press, 460 – 465, 1997.

[39] John McCarthy and Patrick Hayes. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In B. Meltzer and D. Michie (editors), *Machine Intelligence* **IV**: 463 – 502, 1969.

[40] John McCarthy. Circumscription - a Form of Non-monotonic Reasoning. *Artificial Intelligence*, **13**: 27 – 39, 1980.

[41] D. Hugh Mellor. *The Facts of Causation*. Routledge, London and New York, 1995.

[42] D. Hugh Mellor. *Real Time II*. Routledge, London and New York, 1998.

[43] Maurice Pagnucco and Pavlos Peppas. Causality and Minimal Change Demystified. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, Seattle, 125 – 130, 2001.

[44] Pavlos Peppas. *Belief Change and Reasoning about Action. An Axiomatic Approach to Modelling Inert Dynamic Worlds and the Connection to the Logic of Theory Change*, PhD thesis, Department of Computer Science, University of Sydney, 1993.

[45] Pavlos Peppas, Maurice Pagnucco, Mikhail Prokopenko, and Norman Foo. Preferential Semantics for Causal Fixpoints. In Abdul Sattar (editor), *Proceedings of the Tenth Australian Joint Conference on Artificial Intelligence*, Perth, 197 – 206, Australia, 1997.

[46] Pavlos Peppas, Maurice Pagnucco, Mikhail Prokopenko, Abhaya Nayak and Norman Foo. Preferential Semantics for Causal Systems. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, Stockholm, 118 – 123, 1999.

[47] Huw Price. *Time's Arrow and Archimedes' Point*. Oxford University Press, 1996.

[48] Mikhail Prokopenko and Pavlos Peppas. Modelling Inertia in Action Languages (Extended Report). In G. Antoniou, A. K. Ghose and M.Truszczynski (editors), *Inducing and Reasoning with Complex Representations*, Springer Verlag Lecture Notes in AI, Volume 1359, 236 – 249, 1998.

[49] Mikhail Prokopenko. Situated Reasoning in Multi-Agent Systems. In *AAAI Technical Report SS-99-05. The AAAI-99 Spring Symposium on Hybrid Systems and AI*, Stanford, 158 – 163, 1999.

[50] Mikhail Prokopenko, Maurice Pagnucco, Pavlos Peppas, Abhaya Nayak. Causal Propagation Semantics — A Study. In *Proceedings of the 12th Australian Joint Conference on Artificial Intelligence*, Sydney, 378 – 392, 1999.

[51] Mikhail Prokopenko, Maurice Pagnucco, Pavlos Peppas, Abhaya Nayak. A Unifying Semantics for Causal Ramifications. In R. Mizoguchi and J. Slaney (editors), *Proceedings of the Sixth Pacific Rim International Conference on Artificial Intelligence*, Springer Verlag Lecture Notes in AI, Volume 1886, 38 – 49, 2000.

[52] Raymond Reiter. A Logic for Default Theory. *Artificial Intelligence*, **13**: 81 – 132, 1980.

[53] Raymond Reiter. Nonmonotonic Reasoning, *Annual Review of Computer Science*, **2**: 147 – 186, 1987.

[54] Erik Sandewall. *Features and Fluents*. Oxford University Press, 1994.

[55] Erik Sandewall and Yoav Shoham. Non-monotonic Temporal Reasoning. In Dov M. Gabbay, C. J. Hogger, and J. A. Robinson (editors), *Handbook of Logic in Artificial Intelligence and Logic Programming*, Volume 4, Epistemic and Temporal Reasoning, Clarendon Press, Oxford, 1995.

[56] Erik Sandewall. Assessments of Ramification Methods that Use Static Domain Constraints. In L. Aiello, J. Doyle, and S. Shapiro (editors), *Proceedings of the Fifth International Conference on Knowledge Representation and Reasoning*, Morgan-Kaufmann, 1996.

[57] Erik Sandewall. Underlying Semantics for Action and Change with Ramification. In *Linköping Electronic Articles in Computer and Information Science*, Volume 1 (1996): 2, 1996.

[58] Erik Sandewall. Cognitive Robotics Logic and its Metatheory: Features and Fluents Revisited. In *Linköping Electronic Articles in Computer and Information Science*, Volume 3 (1998): 017, 307 – 329, 1998.

[59] Yoav Shoham. Chronological Ignorance: Experiments in Nonmonotonic Temporal Reasoning. *Artificial Intelligence*, **36**: 279 – 331, 1988.

[60] Yoav Shoham. *Reasoning about Change*. MIT Press, Cambridge, Massachusetts, 1988.

[61] Lynn Andrea Stein and Leora Morgenstern. Motivated Action Theory: a Formal Theory of Causal Reasoning. *Artificial Intelligence*, **71**: 1 – 42, 1994.

[62] Michael Thielscher. Computing Ramification by Postprocessing. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, 1994 – 2000, 1995.

[63] Michael Thielscher. Ramification and Causality. *Artificial Intelligence* **89**: 317–364, 1997.

[64] Michael Thielscher. How (Not) To Minimize Events. In *Proceedings of the Sixth International Conference on Knowledge Representation and Reasoning*, Morgan-Kaufmann, 60 – 73, 1998.

[65] Michael Tooley. *Causation. A Realist Approach*. Clarendon Press, Oxford, 1987.

[66] Michael Tooley. *Time, Tense, and Causation*. Clarendon Press, Oxford, 1997.

[67] Hudson Turner. Representing Actions in Default Logic: a Situation Calculus Approach. In *Proceedings of the Third Symposium on Logical Formalizations of Commonsense Reasoning*, Stanford, 1996.

[68] Hudson Turner. A Logic of Universal Causation. *Artificial Intelligence* **113(1-2)**: 87–123, 1999.

[69] Georg Henrik von Wright. On the Logic and Epistemology of the Causal Relation. In E. Sosa and M. Tooley (editors), *Causation*, Oxford University Press, 1993.

[70] Marianne Winslett. Reasoning about Actions Using a Possible Models Approach. In *Proceedings of the Seventh National Artificial Intelligence Conference*, San Mateo, CA., Morgan Kaufmann Publishers, 1988.

[71] Robin J. Wilson. *Introduction to Graph Theory*. Longman Scientific and Technical, 3rd edition, 1985.

[72] Yan Zhang. Specifying causality in action theories: A default logic approach. *Theoretical Computer Science*, Volume 220, 2, 489 – 513, June, 1999.