Pretending to work:
a closed world reasoning formalisation of pretend play

MSc Thesis *(Afstudeerscriptie)*

written by

Lisa Benossi
(born November 1st, 1993 in Lestizza, Italy)

under the supervision of **Prof Dr Francesco Berto** and **Prof Dr Michiel van Lambalgen**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense:	**Members of the Thesis Committee:**
June 28th, 2017	Dr Katrin Schultz
	Dr Luca Incurvati
	Dr Jakub Szymanik

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

Pretend play is often defined as an imaginative play that involves *acting as if*: for example, pretending to work would be analysed as "acting as if one was working". The question on how human beings understand pretence becomes particularly interesting, as soon as one realises that 24 months old children are able to engage in basics forms of pretend play.

In the attempt to clarify what *acting as if* means, the present work deconstructs pretence in terms of simpler reasoning processes, i.e. the ones that children display when they start engaging in pretend plays. This deconstruction is guided by experimental results about imagination, hypothetical reasoning and pretence in early childhood. At the same time, the theoretical analysis of these imaginative phenomena is directed towards a logical formalisation of pretence. I suggest that the logic based on closed world reasoning – i.e. the treating of all the information not currently considered or represented as false – used in this work displays how imagination, subtractive reasoning and pretend play are related.

Since autism is often diagnosed on the basis of a lack of pretence behaviour, part of the work is devoted to the investigation of how children with autism engage in pretend play. The union of psychological, philosophical and logical analysis, presented in this work, models the behaviour of both neuro-typical and children with autism, and generates a novel hypothesis on children's understanding of pretence.

Acknowledgements

First and foremost, I would like to thank my supervisors, Francesco Berto and Michiel van Lambalgen, whose support to my research has exceeded what I could have wished for. Without your enthusiasm, curiosity and extreme intellectual precision, Franz, – always conjoined with some non-standard formalism – I would have never started working on this topic. Moreover, I would like to express my gratitude for your constant support, guidance and for all the discussions which have shaped this thesis. You probably see, better than anyone else, how much they have shaped my metaphysical ideas.

This work wouldn't have been quite the same without the help Michel van Lambalgen. My work on the formalisation of pretend play here presented has hugely benefited from your guidance. Furthermore, I believe it will be apparent how your work on logic and reasoning radically changed my approach to philosophy. More than everything else, I would like to express all my gratitude for having showed me the best side of academic research: the authentic commitment to the research of what is true.

The ideas underlying this work have been presented, *in nuce*, at the "Imagination and Modality Workshop" hosted at the University of Padua: I would like to thank the organisers, Massimiliano Carrara and Vittorio Morato, and the participants, for their kind feedback. Furthermore, I cannot forget to acknowledge the members of my thesis defence, i.e. Katrin Schultz, Luca Incurvati and Jakub Szymanik, for the helpful suggestions they provided and for the vivid intellectual exchange they fostered.

I shall also like to thank my third (unofficial) supervisor, Marco Bacchin: without whose expertise in Logic Programming I would have abandoned the crazy formalism used in this thesis months before grasping its potential. My naive views on mathematical impossibilities and their conceivability have benefited from (late-night) conversations with Levin, whom I wish to thank. The residual naivety of my views on the matter is mainly due to my ignorance. Furthermore, I cannot forget to express my gratitude towards my "sisters", Martina and Frida, tireless proof-readers of my writings. Being unable to mention each one of them, I would like to thank all my mates during these two years in Amsterdam, who have been "who [and what], in the midst of the hell, are not hell". Last but not least, I would like to thank the generous founder of my studies in Amsterdam: to the most passionate workaholic I know, and the person who thought me everything that matters.

Table of Contents

List of Figures

Figure **Page**

Chapter 1

Introduction

Imagination frees us from the tyranny of the present, of the logical, of the "real". It also frees us from the constraints of the now [...]

— David A. Hogue, *Remembering the Future, Imagining the Past*

Human beings are able to *imagine* things that differ from what they are currently perceiving, and situations that contradict what they know to hold true in reality. Therefore, humans can escape from the tyranny of reality by finding shelter in imagination.

The conception of imagination expressed in the quotation from Hogue corresponds to what Kind and Kung [63] call the *transcendent use* of the imagination. I am going to argue, however, that incursions in the realm of imagination do not *only* offer a shelter from the "tyranny of the real". The imagination involved when reading literature, for instance, is not devoid of consequences on our ordinary life: literature *teaches* us how to emotionally react to certain events, or how to deal with situations which we would rarely encounter in real life such as a knight fight or a detective investigation. Furthermore, numerous studies on pretend play suggest that pretence imagining of alternative situations and scenarios instructs children on how it is socially appropriate to act in them or train their abilities to "read intentions", or "simulate other's minds" (see Kavanaugh [61],Lillard [77],Weisberg [121]). Recently, various authors have supported the idea that imagination, in general, can – if grounded in reality – instruct us about the real world, i.e. we acquire useful knowledge through imagining (see Kind and Kung [63] and [64]). Following Kind and Kung, let me call this second usage of imagination *instructive*.

Mimicking a Kantian transcendental argument, one could consider the fact that humans can envisage multiple alternative scenarios as datum, and investigate which faculties and capacities are necessary to explain this fact about human cognition. A first approximation of this procedure for the imagination involved in pretend play will be advanced throughout the present work. The arguments proposed will have a different strength from the ones proposed by Kant, for I will only be able to suggest which processes and abilities may ground the ability to pretend, on the basis of the capacities displayed by children in experimental settings simultaneously to their first episodes of pretend play.

For the moment, however, let me start off providing a less airy characterisation of imagination (section 1.1), and its ties with modality (section 1.2).

1.1 A variety of imaginations

If anything seemed uncontroversial within the vast literature concerning imagination, it would be that the usage of the term "imagination" hides such a multi-layered variety of diverse senses that cannot be captured by a simple taxonomy (see for instance Strawson [109, p. 31], Gendler [35], McGinn [81, p. 595], Kind and Kung [63, p. 3 and ff] and Walton [120, p. 19]). It is therefore practically impossible, within this introduction, to satisfactorily present all the analyses of imagination and its diverse forms. Nonetheless, I shall, at least, hint at some important distinctions. I will attempt to do so presenting some ways of partitioning the possible uses of "imagine":

(1) I imagine that the grass is green.
(2) I imagine a tiger.
(3) I imagine myself skiing (for the first time).
(4) I imagine (towards a contradiction) that Fermat last theorem is false.
(5) I imagine that I didn't come to study in Amsterdam.

(1) states that I represent to myself that something, i.e. the grass is green, is the case. This type of imagining is usually called *propositional* , as opposed to *non-propositional* imagining (see Jackson [54], Walton [120] and Gendler [35]). The latter can be *objectual* as in (2), where to imagine means to stand in a certain relation with the representation of an object, or *active* as in (3). When I imagine skiing, I am simulating how the activity of skiing would look like from my eyes. Furthermore, I would be inclined to say that both (1), (2) and (3) are underlined by the formation of *mental image*, depicting the tiger or what I would perceive skiing. If (2) and (3) are compared with (5), it seems that in imagining that I didn't study in Amsterdam is not purely pictorial, but rather it involves a conceptual dimension, i.e. "conceptually entertaining a possibility" [1]. Similarly to (5), in the case of (4) I appear to conceptually engage with the idea that Fermat's last theorem is false. However, I am generally unable to vividly represent to myself a situation which would make Fermat's last theorem false, while I am quite able to realistically represent myself as studying somewhere else and never coming to Amsterdam. Following Goldman [37], I will say

[1] How do I even know that I could have not come to study in Amsterdam, i.e. that "not coming to study in Amsterdam" was a possibility for me? I have reasonable evidence to think that "I came to study to Amsterdam" is true, but the question as to how I am justified in thinking that it could have not happened seems less straight forward. The subsequent section will be devoted to a brief introduction to some relevant theories.

that (4) is more akin to suppositional or S-imagining, while (3) is similar enough to enactment or E-imagination[2]. *En passant*, notice that the entertaining of E-imagining is characterised by Goldman as likely to generate further mental states and affective responses, while S-imaginings are not. For these reasons, some authors (Doggett and Egan [25]) hypothesised that E-imaginings are involved in pretend play, since they are able to motivate actions.

One could also distinguish mental representations in terms of how they represent. Most cognitive scientists agree that representations can be picture-like or language-like: the former encode information as a map or drawing representing a 3-D space, i.e. with is some minimal level of arbitrariness in that some details are left out of the picture, and both form and size can vary. The latter, instead, seems to be completely arbitrary (see Paivio [86] and Berto and Schoonen [7]). The debate as to whether mental representations are of both kinds, or if one of the two is reducible to the other is the matter of a controversy (see Thomas [111, par. 4.4])[3]. (4) definitely cannot represent like a picture, so if it does, the outcome will be more similar to a linguistic representation. (1)–(3) can represent pictorially, while (5) can represent linguistically, or pictorially with the aid of some conceptual or linguistic labelling (see Berto and Schoonen [7]).

In order to provide a better grasp of which *instructive* use imagination may have, consider again (3). If I actually have to go skiing for the first time, imagining how I could feel skiing may be helpful to make plans on how to move (or not to move), and to make predictions about how difficult it will be for me not to fall (and how to avoid this). Similarly, imagining that I didn't come to study in Amsterdam – as long as my imagining tries to be as realistic as possible – can help me to realise what I couldn't have experienced if I didn't come to Amsterdam.

Finally, in relation with the distinction between (4) and (5), one may wish to distinguish a type of imagining that guides our ideas about possibility from merely supposing. In order to suppose, towards a contradiction, that Fermat's last theorem is false, it suffices to know what it means for Fermat's last theorem to be true or to be false, without "constructing" in one's imagination a situation that would make it false. As a matter of fact, I *cannot* imagine which situation would make Fermat's last theorem false,[4] partly because of mathematical ignorance. However, I can imagine a situation such that "I didn't come to study in Amsterdam" is true. Furthermore, my imagining that I didn't come to Amsterdam suggests me that I could have not come to Amsterdam, i.e. this is a (unrealised) possibility.

[2] The question as to whether (5) will be considered a S-imagining or E-imagining depends on the interpretative stance taken: following Goldman [37], it would seem that because is an imagining *that* it would be a S-imagining, while if we give a primary role to the vividness of the mental representation, I would say that (5) could count as enactment. For sure,“ (5*) I imagine never coming to study in Amsterdam" will count as an E-imagining.

[3] Berto and Schoonen [7] argue that only pictorial representations permit to relate imagination with possibility, if the two are related at all.

[4] Especially if I am not allowed to imagine a bright mathematician claiming that she proved that Fermat's last theorem is false nor imagine myself believing that it is false.

1.2 Imagination, modality and counterfactual reasoning

David Hume famously introduced the maxim according to which "whatever the mind clearly conceives [...] includes the idea of possible existence, or in other words [...] nothing we imagine is absolutely impossible" (see Hume [52, p. 32]). It is the matter of vivid controversy whether Hume's maxim is true and we are justified in using it. Hume's maxim mentions both conceivability and imagination: it is generally accepted that conceivability is a specific kind of imagination, some form of which would guide our intuitions on possibility.

Recently, authors such as Yablo [125] and Chalmers [16] proposed a precise hierarchy of kinds of *conceivability* on the basis of the shared intuition that in some sense conceiving involves an appearance of possibility, i.e. whenever an agent conceives that φ, the agent enjoys something akin to the impression that φ is possible (Yablo [125, p. 5]). According to Chalmers [16], the best guide for possibility is ideal positive conceivability. Ideal conceivability requires justification that cannot be rationally defeated, while with positive conceiving (of φ) Chalmers refers to the imagining of a specific configuration of objects and properties making true φ, or verifying it. More precisely, φ is positively conceivable for an agent if she can imagine a situation verifying φ and s.t. by filling in arbitrary details contradictions do not arise.

Many authors have studied the important role played by imagination in the ability to reason about ways in which things might be or might have been (Byrne [13],Williamson [122],Roese and Olson [98]). Further relating imagination, modality, and counterfactuals, Chalmers [16, p. 10] and [17] suggests to distinguish between two types of conceivability that mirror two ways of thinking about hypothetical possibilities: *epistemically*, as ways in which the world might actually be, and *subjunctively*, as counterfactual ways the world may have been. A sentence φ is primary or epistemically conceivable if one can imagine a situation verifying φ as actual, while φ will be considered secondary or subjunctively conceivable if the subject can imagine a situation verifying φ as counterfactual.

A priori truths would, according to Chalmers, constitute the basis for epistemic possibilities: for all we know *a priori* about the world, water may be H_2O or XYZ and Hesperus may be Phosphorus or not. For secondary possibilities, instead, the denotation of the terms used is fixed with regards to the actual world. This means that it's not secondary conceivable that Hesperus was not Phosphorus, because it is not secondary conceivable that Venus is not Venus and both Hesperus and Phosphorus actually refer to one celestial body (Venus).

According to Williamson [122], the cognitive processes underlying ordinary counterfactual reasoning are also able to handle metaphysical modality. Counterfactual reasoning, according to Williamson, is often generated by imagination – constrained by ordinary perception and background knowledge – in a sort of offline simulation. Williamson [122, esp. chap. 5] and [123] is supported by the results in Byrne [13] and Harris et al. [45] in that counterfactuals seem to be related both with imagination and with causality. The close tie with cognitive capacities implies that humans' imaginative simulations to evaluate counterfactuals are fallible, and we can go wrong. This also means that our intuitions about metaphysical modality are *fallible*, even if we are rather proficient reasoners on counterfactuals and metaphysical possibilities. In summary, Williamson suggests that our sense for metaphysical necessity and possibility is a special case of the ordinary "epistemology of counterfactual thinking" [122, p. 158]. Resulting from his approach, Williamson is able to claim that imaginative evaluation of counterfactuals comes with a conceivability test for possibility (and symmetrically an inconceivability test for impossibility): φ is not conceivable if the counterfactual unfolding of φ robustly leads to a contradiction. However, according to Williamson [122], the distinction between *a priori* and *a posteriori* gets blurred, for imagination is constrained by the agent's general knowledge.

1.3 Where to locate the present work

Within this thesis, I focus on one specific imaginative phenomenon, i.e. the one involved in pretend play, and investigate its connections with counterfactual reasoning. The choice of analysing this specific class of play strongly relies on the intuition that pretending involves imagination. A big share of the philosophical, psychological and cognitive literature on pretence agrees upon the ties of this latter with imagination (with some exceptions, e.g. Ryle [100] and Currie and Ravenscroft [21]). It is sometimes claimed that people can act *as if*, i.e. pretend, without imagining, just as much as they can imagine without acting *as if*. The stance I defend within these pages envisions pretence understanding as a core part of pretend play, even if the former does not involve action, while I still distinguish it from imagination in its broadest sense. On this basis, I will assume that pretending always involves some imagining.

I won't be able to maintain many of the distinctions between types of imagination, for children as young as 24 months are observed freely switching between propositional, objectual and active imaginings. In any case, it is possible to identify an extreme limit of the phenomenon in suppositional imagining: to the best of my knowledge, examples of pretending based on abstract suppositions such as "Let's pretend that Fermat's last theorem is false" in children or adults are not common, except when doing mathematics. It seems plausible that acting as if φ were true requires a good grasp of what it means for φ to be true, and the ability to identify some situation and behaviour which would make true or made true by φ. If this is so, acting as if some complex mathematical (or theoretical) truth were false seems extremely difficult even for an expert in the field. Therefore, the imagination I will consider within this thesis will be closely related with Goldman's E-imaginings, in that it can generate further imaginings and it can motivate action. Finally, studies on pretence understanding in early childhood (see Kavanaugh and Harris [62],Harris [42],Harris et al. [46]) seem to suggest that many pretend plays rely on a pictorial representation, nicely in agreement with the theoretical arguments advanced in Berto and Schoonen [7]. Aside from this, I will attempt to treat the term "imagination" as freely as possible, so as to let the experimental data about pretending guide the discussion about which type of imagination grounds pretend play.

The type of imagination sought will turn out to be similar enough to the one described by Williamson [122]. The imagination described in this thesis, indeed, constitutes the basis for both counterfactual reasoning and pretending. Furthermore, this type of imagination is akin to a form of positive conceivability, in that imagining will always involve the mental (re-)construction of a circumstance or scenario making the imagined situation true. This won't be ideal nor necessarily coherent, though, for children are clearly not perfect reasoners and the circumstances imagined will never be expanded as to constitute a maximally consistent set. On the basis of the literature concerning the development of pretend abilities and of planning in general (see van Lambalgen and Stenning [117]), it appears that human reasoning in these spheres relies on incomplete and partial models, which can be extended and modified at later points of time.

Finally, the imagining displayed in pretend play does often involve alterations of known facts about the world, of causal and physical rules about reality, and inclusion of absent objects or properties. This means that pretending, as I characterise it throughout the thesis, sometimes violates *a posteriori* necessary truths, such as I pretend to be a cat, and more generally metaphysical laws. The extent to which pretend play episodes, in which children engage, affect humans' modal intuitions and conceptions of possibility largely exceeds the scope of this thesis. Nonetheless, I hope that the proposal advanced in these pages will draw increasing attention to the understanding of what is "real" and "pretended" in early childhood. I am convinced that the metaphysical discussion concerning modality and imagination will benefit from a more advanced theory of imagination. The study of the manner in which children learn to imagine pretended stipulations and to reason counterfactually, indeed, might reveal something about how humans think of (counterfactual) possibilities. In addition, I hope that the formalisations of imagination, counterfactual reasoning and pretend play advanced within this writing will display that imagination and logical reasoning are not mutually exclusive spheres[5].

1.4 Structure of the thesis

The focus, from now on, will be gently shifted towards pretend play and the imagination employed in it. Therefore, henceforth, imagination will generally refer to the one underlying pretend play, whenever it's not specified otherwise.

[5]The logical framework used in this thesis is a logic for planning, developed in van Lambalgen and Stenning [117] and van Lambalgen and Hamm [115]. The choice of a logic for planning to formalise imagination seems justified, in the case of pretend play, by the consideration that pretence behaviour and understanding are deeply intertwined with the planning of an appropriate response. More in general, the characterisation of imagination as an offline simulation or reconstruction (see Currie and Ravenscroft [21], Goldman [37]) reveals a tight connection with the offline simulations employed in planning. Arguments supporting the cognitive affinity of imagination and planning can be found in van Leeuwen [118], who characterises imagination on the basis of forward models (Frith [32], Schacter et al. [102], and Addis et al. [1]). For a defence of the fundamental role of planning in human cognition, see van Lambalgen and Stenning [117].

The introduction to pretend play will start, in chapter 2, from the exposition of two relevant theories of pretence. By starting with an analysis of the theories proposed by Nichols and Stich [85], and by P. L. Harris [43], I will provide the reader with a grasp of the common features to many different theories, while simultaneously introducing different possible approaches to pretence. Not surprisingly, both Nichols and Stich and P. Harris underline an important connection between pretence and counterfactual reasoning. The differences in conceiving this relation will motivate an incursion into the cognitive and developmental data concerning counterfactual reasoning in early childhood. On the basis of the latter, I will sketch a plausible hierarchy of senses in which "counterfactual" reasoning develops in childhood and adulthood.

Since autism is often diagnosed on the basis of a lack of pretend play, explaining how children with autism differ from normally developing children in pretence production and understanding has become an essential criterion for a theory of pretence. Therefore, I will embark on the attempt to display that the dissociation hypothesis formulated in van Lambalgen and Stenning [117],Smid [107],van Lambalgen and Smid [116] and Pijnacker et al. [92] can explain children with autism's performance on pretend tasks without relying on *ad hoc* hypotheses. The *dissociation* hypothesis, roughly, suggests that the capacity to handle exceptions to a rule is selectively impaired in autism, while reasoning about rules themselves is intact. According to van Lambalgen and Stenning [117],Pijnacker et al. [92] these two types of reasoning are normally deeply intertwined in the mental processes underlying planning and discourse interpretation, and they appear dissociated in autism (see also Smid [107]). The second chapter will find a close with an overview of children with autism' behaviour on pretend tasks, and with the introduction of the *dissociation hypothesis*.

The third and fourth chapters will advance a theory of pretend play, constructed on the basis of the criticism I will advance to Nichols and Stich' and to P. Harris' theories in 2.4.2 and on the dissociation hypothesis. Chapter 3 will mainly focus on the theory of pretence, while chapter 4 will introduce the logical formalisations of the theoretical notions in chapter 3. They can be read independently, even though I conceive them as complementary, in that only together they provide a complete picture of the theory of pretend play I advance. Furthermore, their structure is almost symmetric: for most sections in chapter three the reader will find a corresponding section describing the formalisation I employ.

Chapter 3 will focus on the theoretical description of how pretend play and "counterfactual reasoning" stem from *counterfactual imagination*, i.e. any mental representation which differs from the current representation of reality. On the basis of the hierarchy of types of "counterfactual" reasoning proposed in chapter 2, I will distinguish *subtractive reasoning*, i.e. reasoning about what would happen if an event e, happening in reality, did not happen, from full-fledged *counterfactual reasoning*. The latter refers to reasoning about two mental "files" or representation, one of which stands for reality, and the other is obtained by altering the representation of reality as little as possible[6]. The studies in Perner et al. [88] and Cristi-Vargas et al. [20] suggest that

[6]Therefore, both subtractive reasoning and full-fledged counterfactual reasoning involve two mental representa-

below age of 12 humans are imperfect counterfactual reasoning, in that basic conditional rules guide counterfactual reasoning. This entails that the ability to alter as little as possible the representation of reality is acquired in a rather long time span. In the formal counterpart of this section, i.e. section 4.4, I will display, with the aid of logical tools, why counterfactual reasoning in its full-fledged sense is cognitively requiring, and how the reliance on basic conditional rules facilitates reasoning about what would have happened if an event didn't take place.

Chapter 3 will find its conclusion in an highly theoretical explanation of how pretence works in both neuro-typical and autistic children. The result will be mirrored in chapter 4, where I propose a full logical formalisation of one type of pretend play, i.e. object substitution. Since the *dissociation hypothesis* finds a natural logical formalisation, the formalisation I will propose of object substitution in children with autism generates some novel explanation of the observed stereotypical behaviour of children with autism in pretence episodes. For reasons of space, a complete formalisation is proposed only for one type of pretend play, but on the basis of the definitions advanced in chapter 4, a formalisation of all types of pretending can be achieved.

The thesis will be concluded with an overview of what has been achieved, and which elements can be further expanded and improved in further work. In chapter 5 I will also hint at some interesting directions that have exceeded the scope of the present writing, but which would constitute interesting research directions. I will conclude the original part of the thesis by suggesting which relevance it may have, if compared with past and future studies of imagination, pretence and counterfactual reasoning.

The reader will find, after the conclusion (chap. 5), two appendices. Appendix A is meant to constitute an aid for the reader, which may not be familiar with the logic I use in chapter 4. Most of appendix A introduces the work of authors from whom I took inspiration to achieve the desired formalisation. The extension of pre-existing logics I advance and employ within the work is also introduced in appendix A.

Appendix B briefly presents a modification of a pretence understanding task, presented in Kavanaugh and Harris [62]: during the investigation of pretence understanding and production in children in autism, I have been surprised by the paucity of experiments that test whether children with autism understand pretence transformations as real or not. On the basis of the results in the reality vs appearance distinction [29], and of the results in Bigham [9], one could hypothesise that children with autism are more likely to err in questions as to whether a certain pretend property or transformation is "real" or "pretended". In line with the theoretical account I propose in chapter 3, I suggest that it is possible to test the reality vs pretence distinction, without relying a certain interpretation of the word "real" (see Bunce and Harris [11]).

tions, one about reality, and the other obtain through some alteration of the reality file. Furthermore, both types of reasoning require the ability to compare the reality file with the imagined alteration. Subtractive reasoning, however, mustn't modify reality as little as possible, since it works on the basis of simple conditional rules of the form "If φ, then ψ".

CHAPTER 2

Pretend Play

One of the first manifest examples of imagination in children's development is pretending. The aim of this chapter is to understand pretence in its relationship with imagination and reasoning against known facts. In order to achieve this aim, I am going to introduce two successful theories of pretend play, and discuss the available empirical data about pretending and counterfactual reasoning. Furthermore, since an abnormal behaviour in pretend play is one of the earliest symptoms of autism, many studies have been conduced to understand how children with autism perform in pretence tasks. Therefore, a necessary requisite for a satisfying theory of pretence is to explain the differences in autistic children's pretending relative to neurotypical children.

2.1 Introduction

Across the various theories of pretend play that have been proposed in the literature, certain recurrent elements can be identified as necessary [1]:

1. children must act *as if*, i.e. engage in a non-literal simulated activity;
2. pretend actions must involve both actual and non-actual properties of objects;
3. pretence is a mental activity intentionally directed towards something.

On the basis of these three core elements of pretence, I am going to introduce two theories of pretend play: the one proposed by Nichols and Stich in Nichols and Stich [85], and the one

[1] See Leslie [71],Lillard [76],Kavanaugh [61],Lillard [76],Lillard [77],Carruthers [14],Carruthers and Smith [15],Nichols and Stich [85],Harris [43],Harris et al. [44] *et cetera*. Notice that the definition proposed aims at applying to non-human animals as well.

advanced by Paul Harris (see Harris [43, 42] *et cetera*). The choice of focusing on these two models of pretend play is motivated by the high level of refinement and the explanatory power enjoyed by both theories. Furthermore, the two approaches can be viewed as advanced developments of two extremely successful families of theories about "mind-reading"[2]: the Theory-Theory and the Simulation Theory. However, the particular attention to Nichols and Stich's and P.L. Harris' work won't be an impediment to the analysis of a wider spectrum of cognitive, psychological and philosophical literature on imagination and pretend play.

The chapter is structured as follows: section 2.2 introduces two families of theories concerning mind-reading: the Theory Theory and the Simulation Theory. A brief introduction to these general stances on mind-reading is going to gently introduce the reader into the two theories of pretence that I will analyse more in details.

Section 2.3 presents the theory of pretend play proposed by Nichols and Stich. Furthermore, I display how their theory of pretend play relates with their general stance on mind-reading abilities. I finally propose few critical observations concerning Nichols and Stich's model of pretend play.

The fourth section (sec. 2.4) introduces an alternative theory of pretend play, the one advanced by Paul Harris. According to Harris [43], the ability to engage in counterfactual reasoning is essentially different from the capacity to imagine displayed in pretend play by 2 years old (or older) children[3]. Some thoughts on Harris' model of pretence are introduced in section 2.4.1. Finally, in section 2.4.2, a comparison between the two models is proposed, to explicitly discern the merits and the differences between the two accounts.

The discussion of the differences between Harris' [43] and Nichols and Stich's [85] accounts of pretence is going to lead me to hypothesise that a slightly different understanding of the term counterfactual underlines the two works. Therefore, in section 2.5 I analyse the type of "counterfactual" reasoning that 2-years old children are able to do. This overview aims at establishing which sense of "counterfactual" is plausibly intended by both Harris, and Nichols and Stich. Therefore, a survey of the studies in cognitive and developmental psychology concerning counterfactual reasoning in childhood is undertaken.

Finally, section 2.6 presents some of the most important studies on children's pretence. Many relevant theoretical distinctions – which will be used and referred to in later parts of the work– are going to be introduced. Furthermore, since Autism Spectrum Disorder (ASD) is generally diagnosed and characterised in terms, among other things, of difficulties or impairments in imagination and pretence, I briefly present some cornerstones of the study of pretend play within the *ASD*.

[2]Notice that I do not imply that neither Harris nor Nichols and Stich claim that pretend play necessarily involves mind-reading. In fact, they claim the opposite, but the claim can be understood with diverse strengths depending on the sense attributed to the term "mind-reading". Further explanations will be provided at in section 2.2.

[3]More precisely, the ability to pretend typically emerges between 18 and 24 months of age (see Leslie [72]).

2.2 A brief introduction to theories of "mind-reading"

The term "mind-reading" generally refers to the ability to understand and intuit human minds. Examples of this ability pervade our everyday lives: for instance, a child that affirms that she is avoiding a certain behaviour because her mother would otherwise be upset displays *in nuce* the capacity to "read" other people's feeling and mental states. Within this work, I use the term as a general label for a wide spectrum of human skills rather than a basic ability itself. Notice, *en passant*, that it is often claimed that pretence is, generally, the first sign of mind-reading abilities in children's development (see for instance Leslie [71], Leslie [72, p.3], *contra* this Nichols and Stich [85] and Harris [43]).

The study of mind-reading has been the subject of intensive discussion in developmental psychology. One of the most influential theories concerning mind-reading has been advanced by Alan Leslie, who described the abilities that underlie pretence as requiring *meta-representations*. In extreme synthesis, Leslie proposed that humans own a specific mental module that takes care of reasoning about other minds (see Leslie [71],Leslie [72], van Lambalgen and Stenning [117, p.248-49]). The module underlying mind-reading has also been called *theory of mind mechanism* (ToMM). Literal and transparent descriptions of the world, i.e. *primary representations*, have a "decoupled" counterpart, i.e. for any primary representation there can be an opaque version of it. The decoupling allows representations to be embedded in larger representational structures (e.g. attitudes). This complex relational structure is called by Leslie a "M-representation". According to Leslie [72], the ToMM employs M-representations. Following this theory, people with autism have a defective or impaired module for reasoning about other minds.
The ToMM relates to, even though it doesn't necessarily overlap with, the so-called Theory Theory (TT). The latter suggests that prediction, explanation and interpretation of intentional states such as beliefs and desires exploits an internally represented theory or knowledge structure, also known as "folk psychology". The ToMM hypothesis and the TT have found support for their claims from the False Belief Task (FBT), which seem to display that children before the age of 3 and children with autism have issues with the theory of mind (ToMM) or with folk psychology. As far as children below the age of 3 are concerned, the ToMM hypothesis claims that the ToMM is not yet sufficiently developed. For children with autism, the ToMM hypothesises that the ToMM is impaired or deficient.

An alternative popular explanation of how humans succeed in guessing other people's feelings and mental states is the *simulation theory*. The simulation theory (ST) explains mind-reading abilities through an "off-line" process which employs imagination, pretence and perspective taking to "simulate" the mental states of someone else (see Shanton and Goldman [106],Gordon [38], Heal [48], Harris [40, 43]). Intuitively, and roughly, a child reads someone else (i.e. the target) mental states by simulating what she would have believed or thought if she had the information available to the target. According to ST, mind-reading abilities can be distinguished in low-level and high-level abilities. A feeling for the distinction can be grasped through an example. Let's consider the automatic process of attributing emotions on the basis of facial expressions as an instance of low-level mind-reading (see Gallese et al. [34], Rizzolatti and Craighero [97]). The high-level mind-reading refers to more complex processes that tend to involve propositional attitudes (see Shanton and Goldman [106]).

Intermediate theories have been proposed, as well as alternative accounts of mind-reading. In any case, for the present introductory aims a brief sketch of these two families of theories should be sufficient to understand the fundamental structure of Nichols and Stich's model. Further considerations on the virtues and limits of these approach will be introduced in due course.

2.3 Mind-reading and pretending: Nichols and Stich

Nichols and Stich's attempt at producing an unified and as complete as possible account of human mind-reading abilities starts off with an analysis of pretence. In "Mindreading", they propose a cognitive architecture of pretence play envisaging three kinds of functionally different representational states: beliefs, desires and "possible worlds"[4]. In Nichols and Stich's ([85]) terms, a "possible world" is a representation token, whose functional role is distinct from the one of beliefs and desires. "Possible worlds" represent how the world would be like, if a given set of assumption - whose content one may neither believe to be true nor desire to be true- were the case.

An episode of pretend play, according to Nichols and Stich's account, is initiated by a pretence premise, which is either produced by the pretender herself or which must be reconstructed when someone else initiates the pretence (see Nichols and Stich [85, p.24]).

The pretence premise is accompanied by the *inferential elaboration*: on the basis of the pretence premise, perceptual information, background knowledge, memory *et cetera*, the pretender infers the appropriate consequences of the given assumption.

At a highly theoretical level, Nichols and Stich's theory affirms that pretending is generated by a pretence premise placed in the, so-called, the Possible World Box. The terms possible world and possible world box (henceforth PWB), in Nichols and Stich's writings, diverges from the term's use in the philosophical literature. It possible to understand the PWB as a mental file or

[4]Nichols and Stich's model is intended to be able to represent descriptions of situations that would logically or metaphysically be labelled as *impossible*. This does not include obvious contradictions, but representations that either contradict some general rule about reality or premises that can be discovered to be contradictory through a long and/or complex reasoning process. I will say something about the imagination of impossibilities – both logically and metaphysically speaking – in chapter 3. For the moment, it shall suffice to notice that is reasonable to assume that human beings may imagine a situation that is not possible in the Lewisian or possible world semantics' sense. A clear example would be the construction of a highly incomplete situation for a premise. Notice that this representation cannot be technically be labelled possible, even though it is obviously possible for us to entertain it.

Furthermore, concerning the use of *reality*, henceforth, and throughout the whole work I will use *reality* as a shorthand for the agent's representation of reality. Similarly, I will use the term *actual* to denote whatever is the case in the agent's current representation or model of the world. The usage of "reality" and "actual", therefore, does not entail anything about how the world really is, its essential or metaphysical status, nor that the agent's representation corresponds to what is represented set of representations separated from the representation of reality and describing a way in which the world would be, given a certain set of assumptions.

The role played by the PWB, in Nichols and Stich's cognitive architecture, is to construct an "increasingly detailed description" (see [85, p.29]) of how the world would be if the pretence premise were true. The representation of a situations that makes true the pretence premise is constructed on the basis of our inference patterns, the very same ones that are used in the formation of beliefs. However, insofar as the PWB only contains pretence premises of the sort "This [banana] is a phone", the inferential mechanisms won't be able to construct a very detailed representation. Therefore, the construction of the possible world (making true the pretence premise) relies on the actual beliefs of the agent, as long as they are consistent with it. Nichols and Stich's boxology calls this process "UpDater". According to their model, the UpDater is a sub-part of the inferential mechanisms, which allows us to import as much knowledge and beliefs as possible from the real world to the imagined situation. A powerful metaphor capturing the intuitive role played by the UpDater is offered by Nichols and Stich themselves (see [85, p.32]): one may think of the UpDater as the filter selecting which elements of someone's knowledge and beliefs are allowed to be added to the pretence premise.

The UpDater functions in pretend play and imagination just in the same way in which it would do in belief revision[5]. This ensures that the import of real beliefs to the PWB creates a coherent representation[6]. Admittedly, the exact manner in which belief revision works is matter of extensive debates. Even though Nichols and Stich acknowledge that their model is completely silent on how update works, one may argue that leaving such a central part of the theory undetermined weakens the explanatory power of the theory itself .

Nichols and Stich further assume that the pretender's beliefs include packets of representations, whose content serves as a paradigm describing how certain situations typically unfold. These scripts or paradigms serve to constrain how a possible world concerning a pretence premise φ is constructed. Scripts play an essential role in guiding the imagination and description of a certain situation, as it can be seen in children's reliance on them to describe familiar (temporally extended) activities such as going to eat at a restaurant or baking cookies (see Myles-Worsley et al. [83] and Friedman [30, pp. 91-2] and Nelson [84]). However important, it is clear that the elaboration of scripts does not suffice to exhaust the creativity displayed in imagination and pretend play. Hence, Nichols and Stich allow script's constraints to be modified or reverted by the pretenders' choices and preferences. Within the cognitive architecture proposed by Nichols and

[5]Nichols and Stich present one important difference between standard belief revision and import of knowledge and beliefs in the case of imagination of pretence. In the case belief revision, one may argue, pre-existing beliefs may be modified on the basis of an update (even though some authors have argued that this rarely happens on the basis of a *conservative* tendency, for instance see Harman [39]). For the generation of a possible world representing what would be the case if a pretence premise were true, instead, the pretence premise cannot be removed even if it contradicts some general rule about the world.

[6]Notice, *en passant*, that the manner in which Nichols and Stich characterise descriptions in the PWB entails that they endorse the view that propositional imaginings are belief-like. (see Nichols and Stich [85, p. 32] and Langland-Hassan [69, section 5]).

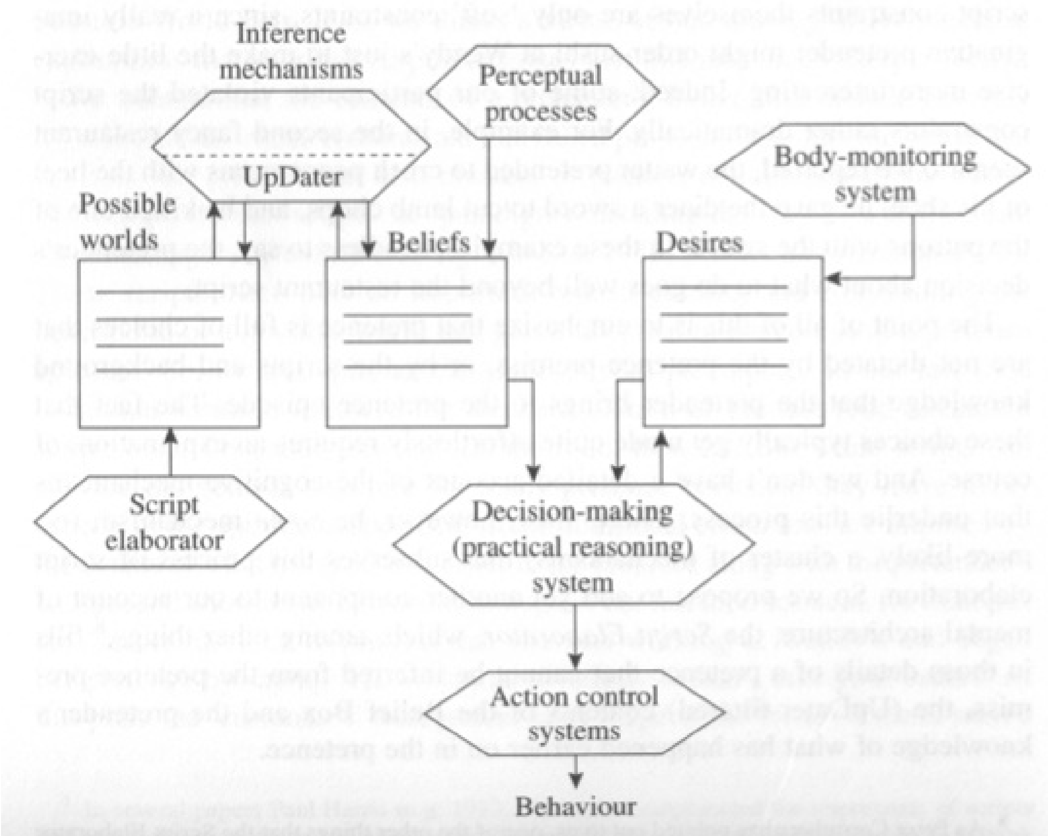


Figure 2.1: Nichols and Stich's model of pretend play (from Nichols and Stich [85, p.31])

Stich, the scripts and creative constraints on pretence are generated by a *Script Elaborator*.

Following Nichols and Stich [85], the description heretofore proposed may be taken as a theory of imagination, while it is not sufficient to explain pretence. Nichols and Stich, indeed, affirm that a crucial element of pretend play is behaviour. If Nichols and Stich's story terminated here, they would have failed in explaining how human beings are able to pretend that φ were the case without literally behaving as if φ were true. They don't, and analyse pretence in terms of behaviour and desire. They analyse pretence as "behaving in a way that is similar to the way some character or object would behave in the possible world whose description is contained in the Possible World Box" (see Nichols and Stich [85, p. 37]). Furthermore, because behaviour is tightly linked to desire, they hypothesise that pretence behaviour stems from the *desire* to behave more or less as she would do if the pretence premise were true.

Finally, the cognitive architecture proposed in Nichols and Stich [85] assumes that beliefs and representations can access the content of the PWB. This is achieved through the import of a conditional rule of the form

If φ were true, then it would (or might) be the case that $\psi_1, \ldots, \psi_n$

where φ is the pretence premise and the ψ_is are the representations in the PWB following from assuming φ. One example of the conditional beliefs deriving from a pretend play could be "If I were a train, I would emit sounds like 'choo-choo'". These conditional beliefs are not *about* pretence, in the sense of being a meta-representation of what pretending to be a train means. Nichols and Stich envisage these conditional beliefs, generated by pretending as the source for hypothetical reasoning in general (see Nichols and Stich [85, p.63-4]).

2.3.1 Some critical remarks

The formalisations of pretend play and imagination I will propose in chapter 3 and chapter 4 have been inspired and motivated by Nichols and Stich's work on pretend play. Therefore, it won't be hard to find similarities both in terms of motivation and of distinctions drawn. However, there are three essential matters on which I am convinced that their model could be improved. I will start by exposing these issues, and I will try to motivate them throughout the chapter. The whole work can be understood as an attempt to improve Nichols and Stich's account on the basis of these "criticisms", and as the result of a critical comparison between the theory proposed by Nichols and Stich, and the one advanced by P. L. Harris.

The first criticism that may be directed against Nichols and Stich's theory of pretend play has already been mentioned in section 2.3. As previously mentioned, the manner in which belief revision and import of real world beliefs and knowledge into the PWB is left completely undefined. Therefore, Nichols and Stich's account is highly vague on how young children acquire the ability to import real world knowledge into imagined situations. Furthermore, by leaving the dynamics of the import unspecified, the theory of pretence proposed in "Mindreading" is unable to answer questions on how this import is executed, and as to whether difficulties in pretending may be due to impairments with it[7]. In fact, the formalisation of pretend play, which I will advance, is able to explain the specific abilities and difficulties of autistic subjects in pretending on the basis of an explicit model of how real beliefs are imported into the pretend representation.

[7]A related criticism to Nichols and Stich's model has been advanced (see Langland-Hassan [69, sect. 5]) with regards to the behaviour of the Script Elaborator. Langland-Hassan argues that Nichols and Stich's suggestion that propositional imaginings are belief-like is highly dubious. As Langland-Hassan [69] says "[w]e are left with a theory that says: imaginings are belief-like ... except for the many ordinary circumstances in which they are not". The solution proposed by Langland-Hassan is that our imaginings can deviate from the scripts and general unfolding of a situation by intentional intervention. This takes the form of a cyclical processing in which lateral processing if "fed" with deviations from the script, which reflect the agent's desire to make things more funny or unusual.

> Guiding Chosen imagining, in its more freewheeling instances, then becomes a kind of cyclical activity, during which new and sometimes unusual premises are "fed" to a lateral algorithm at varying intervals. The output of the lateral "inferential" activity can then, at different intervals, be recombined with a novel element contributed by one's intentions to begin the lateral processing anew (it is because of this recombination that I am calling the process "cyclical"). This allows the imaginative episode, as a whole, to both be constrained (by the lateral algorithm) and to freely diverge from anything one would have inferred from the initial premise alone [...]

The objection raised by Langland-Hassan hasn't been mentioned among the three main criticism that I pose to Nichols and Stich [85] account, however the formalisations proposed in chapter 4 will allow for this sort of cyclical activity on the agent's imagining and pretending. For more details about the formalisation(s) I propose, the reader is directed to chapter 4.

As a second objection to Nichols and Stich's theory, I will attempt at arguing that they seem to conflate pretence understanding and imagination. Since they model pretend play only as the production of a specific behaviour, they appear to assume that pretence understanding is just imagining. However plausible this assumption may seem, Jarrold [56], Kavanaugh and Harris [62], Jarrold et al. [58] display that children with autism are capable of understanding pretence transformations. At the same time, autism is generally diagnosed on the basis of impairments or deficits in imagination. Therefore, I suggest that Nichols and Stich would have to amend their theory to display how children with autism can understand pretend play, and to display how they envisage imagination works for autistic subjects. On this basis, the formalisations proposed in chapter 3 and in chapter 4 will be able to distinguish between imagination and pretence understanding.

The third criticism is better conceived as a starting point for refinement, with regards to Nichols and Stich's account of autism. The account of pretence, presented in section 2.3, suggests that pretence is essentially captured by desire, behaviour and imagination. Moreover, people with autism display difficulties with pretending. On the basis of these premises, one might hypothesise that, in Nichols and Stich's theory, the cause for these difficulties are caused by either these elements – behaviour, desire or imagination – or by the connections between them. Nichols and Stich, indeed, suggest that the difficulties with pretending displayed by people with autism are explained by some damage in the Possible World Box or in its ties with the planner. Hence, they argue *contra* the theory of mind deficit hypothesis, which claims that autistic people lack or are impaired in their ability to read other minds. While I fully endorse Nichols and Stich's arguments against the theory of mind deficit hypothesis (see Nichols and Stich [85, pp. 128-131]), I will propose some more precise hypotheses on how to characterise the interaction between imagination – i.e. PWB– and planning might work in the case of children with autism.

Before devoting our attention to Paul L. Harris' theory of pretence, I would like to underline the tight connection between pretending and engaging in some form of subjunctive or counterfactual reasoning in Nichols and Stich [85]. The result of pretending, according to the latter theory, is the import of conditional beliefs with subjunctive form from the PWB to the Belief Box. Furthermore, the model describes imagination essentially as the construction of some counterfactual situation – i.e. a situation where something that is false (respectively true) in the actual world is imagined to be true (respectively, false). Naively, one may think that this suggests that counterfactual reasoning and imagination are tightly connected, and that children's ability to pretend is a plausible symptom of their ability to engage in counterfactual or subtractive reasoning. Since Paul Harris stresses a difference between pretending and counterfactual reasoning, it will be particularly interesting to compare the two theories in these regards.

2.4 A distinction between pretence and counterfactuals: P. L. Harris

P. L. Harris has provided a huge contribution to the study of pretend play and, more generally, of developmental psychology. Harris [43] extensively describes how imagination, pretending and counterfactual reasoning are developed in early childhood. Furthermore, he advances an evolutionary story of how the ability to imagine something different from reality might have enjoyed so much success and how it might have developed.

Let me start off by introducing Harris' theory of pretend play. I will be able to devote limited space to the discussion of many details of this model, therefore I invite the avid readers to the works of Harris himself (e.g. Harris [41], Harris et al. [45], Kavanaugh and Harris [62], Harris [40], Leevers and Harris [70], Harris [43], Harris et al. [44] *et cetera*).

Since pretending is an important social activity, in that it is generally initiated and guided by caretakers or older mates, and because it teaches children the typically accepted behaviour in certain circumstances, Harris focuses especially on pretend play that are initiated by someone else. According to Harris [43], children by the age of 2 years are able to understand pretend actions carried by a caretaker adopting a pretence stance towards a play episode[8]. This stance can be generally understood in contrast with the literal processing of the immediate environment.

A child that is elicited by a caretaker to "pour some tea" from an empty teapot to a doll is supposed to be able to stop "scanning the immediate environment for situations that literally fit the utterances being produced or ways to comply literally with the requests that are being made" (see Harris [43, pp. 22-23]). If the child sought for literally fitting consequences of the actions, she would be bewildered by the fact that pouring generally involves the transfer of a liquid. Toddlers don't appear confused by these kinds of actions, and to explain how they can adopt a pretence stance it is necessary to introduce a second element. Together with the interruption of the literal processing, they *construct a situation* matching the actions and utterances of their partners through pretence.

Furthermore, the *pretence premise(s)* are inferred from the actions and utterances of their partners, and they are treated as mental "flags" or "reminders". For instance, a child that sees her mother taking up a banana, holding it at her ear and saying "Hello Daddy, how are you doing?", will firstly construct a pretence situation that fits the actions and utterance, i.e. one in which the banana is a phone. Furthermore, the child will encode the stipulation "for the duration of this play episode, the banana is a phone". Within the pretend play episode, the tags are assumed to be retrieved and used to extend the pretend representation and to interpret further actions or utterances. It is important to notice that the pretence stipulations have limited validity in

[8]Children below age of two have difficulties in understanding which effect a pretend action causes. See Harris et al. [46],Jarrold et al. [58], Kavanaugh and Harris [62], Harris et al. [44]. It extremely interesting to notice that around the beginning of the second year of life children both become able to identify the results of a pretended action (using simple heuristics) and start engaging in pretend play.

time: at the end of the pretence episode they are abandoned and the child returns to a literal interpretation of the world. Furthermore, according to Harris [43], by consulting the reminders, a child is able to describe or pick the appropriate picture describing the events happening in the pretend play. This allows Harris to explain how children are able to succeed in pretence understanding tasks. In experimental conditions, when an experimenter, at the end of a pretend episode, asks a child what happened[9], the response reflects what happened in the pretence rather than what the child literally saw.

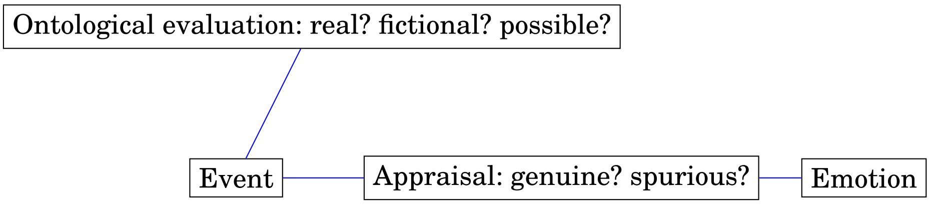
A further element of pretending is *unfolding a causal chain*, which can roughly be assimilated to the inferential mechanism in Nichols and Stich [85]. According to Harris [43], children's understanding of causal connections in pretend play can be understood as a "routine by-product of a simple processing rule" (see Harris [43, p. 23]). Assuming that the entities or substances stipulated by the pretend actions are subject to the same causal rules as their actual equivalents, it is easy to see how causality is conceived in pretence. For instance, in Kavanaugh and Harris [62], the experimenter enacts a pretend action (e.g. pouring tea from an empty teapot) on an animal puppet. The pretend stipulation in this case would be that the tea-pot is full of tea (or some liquid), and on the basis of this flagged information, the child assumes that the pretend liquid behaves like real liquid and hence the puppet will get wet. Harris hypothesises that once these effects have been imagined, they are also added to the mental representation of pretence as flags in the same way as the pretence premise(s).

It should be easy to understand that Harris' theory belongs to the simulation theory's family. As a side note, it is interesting to notice that Harris' general account of imagination offers a theory of how humans can be emotionally affected by an imagined or pretended event, figure or action without relation to the assessment of its real status (i.e. reality vs appearance, reality vs pretence etc). A pictorial representation of this idea is presented in *figure 2.2*. Even though I won't dwell upon the experimental results motivating this theory, the distinction between "ontological" assessment of an imagined event, figure or entity and its emotional assessment is crucial. The emotional involvement, e.g. fear or excitement, created by imagining something and more generally by literature and artistic expressions shouldn't be conflated with the belief that the imagined situation is real. For more details concerning the developmental results, the reader is invited to consult Harris [43].

The procedure for pretence understanding hitherto described may appear similar enough to counterfactual reasoning, insofar as "counterfactual" is understood as broadly as to include the reasoning processes about some situation that differs from the agent's actual representation of the environment. Notice that translating pretence premise into counterfactual premise and pretend consequences into counterfactual consequences, one would get that children just by

[9] For instance, in Kavanaugh and Harris [62], (experiment 2), one of the conditions was showing a puppet animal, e.g. a duck, and enacting a pretend transformation on it, for instance pretending to pour tea on it. Children were shown three pictures: one representing the animal at the beginning of the pretend play, one with the correct transformed choice, and one with an irrelevant transformed choice. They were asked to point at the picture representing how the duck looked like after the experimenter's actions.

Figure 2.2: Flow diagram emphasising the distinction between two forms of assessments: ontological evaluation and appraisal, from Harris [43]



pretending engage in some sort of counterfactual reasoning. At the end, they are able to reason from a stipulation to infer its causal outcome(s).

Harris [43, chap. 6] argues that imagination and pretending differ from thinking about how the world might have been, i.e. what I will call until further notice counterfactual reasoning. Both imagination and counterfactuals belong to suppositional thinking, but, according to Harris, the imagination involved in pretending is not necessarily directed towards reality. Children, the author argues, do not set up a contrast between an imaginary event and a real one. Pretence is rather, for Harris, self-contained, and its events are not conceived as departures from real events. This would explain why pretence may turn on a situation that coincides with reality (see Harris [43, pp. 123-25]).

2.4.1 Some remarks on Harris' model of pretence

The present work has been inspired by P. L. Harris' work just as much as from Nichols and Stich's theory. Before devoting some attention to a comparison between the two, which will lead me to discuss various results in developmental and cognitive psychology about counterfactual reasoning, I believe that it might be helpful to introduce straight away one point on which I essentially disagree with Harris.

In chapter 3.5.4, I argue that an appropriate production of pretence behaviour requires some sort of *comparison* between what's going on in the pretence and reality. Therefore, I reject the claim, in Harris [43], following which pretend events do not require a contrast with reality. In fact, I suggest that it may be impossible to explain the difference between carrying out an action and pretending to carry out the same action without a comparison between what is really true and what is pretended.

Consider the following example: I am working on my master's thesis right now. If I wanted to pretend to work on my master's thesis, I wouldn't act just in the same way as I am doing right now. I would presumably exaggerate certain actions, while substituting the action of actually typing with the gesture of moving my fingers above the lap-top's keys without actually touching them or without thinking about what I should be typing. Even though the most general rules of causality may remain the actual ones in pretending, the actions actually carried out are different. Pretenders know that pretended transformations and actions have different consequences from real actions. For example, if I work on my master's thesis for 8 hours, one could expect me to have written or read a lot of pages, or having thought about very important issues. On the contrary, if I pretend to work for 8 hours, none of these results could be expected.

Imagining a situation is self-contained and requires no comparison with reality. One can imagine all sorts of things without even thinking about what's happening in reality. It may easily happen that imagining distracts the agent from the literal interpretation of her immediate environment, and that hence she does not know anymore what's going on in reality. Differently, in order to pretend, an agent also needs to keep track of what is happening in reality, in order to integrate novelties and events into the pretence settings. Furthermore, to act upon a pretence stipulation, it is necessary to ensure that an action is not literally carried out as if the pretence premise(s) were true, if acting as if the pretence premise were true were to lead to some impediment or contradiction. In order to do so, I argue, the parallel consideration of the reality representation and of the pretence representations is needed. A full fledged argument for this consideration will be constructed throughout chapters 3 and 4.

2.4.2 A comparison between Harris' and Nichols and Stich's models of pretence

At a highly abstract level, the models proposed by Nichols and Stich [85] and Harris [43] agree on the most essential elements of pretence. Leaving aside the objections already advanced in section 2.3.1 and section 2.4.1, and at a less abstract level, however, there are important differences in terms of how the imagination involved in pretend play relates to counterfactual reasoning.

2.4.2.1 An important distinction: counterfactual imagination and subtractive reasoning

Before diving into a comparison between Harris' and Nichols and Stich's theories of pretence, let me dwell on a terminological (and theoretical) distinction, which will be used in the rest of the work.

Various forms of imagination are involved in pretence: throughout the present work I will attempt to show that the capacity to imagine that *something is the case* is the fundamental ability underlying pretence. I will use the term *counterfactual imagination* to denote this extremely simple and basic human ability to deviate from the literal representation of reality within one's imagination. For instance, a child pretending that the (empty) tea-pot is full is deviating from the literal, faithful representation of reality, and she is imagining something that differs from reality. This counterfactual imagination can extend, shrink, alter, revert what the child knows or believes about reality.

A more specific and complex form of counterfactual imagining displayed in pretence is the capacity to imagine that one specific event did not happen and to compute how the world would have been if the event did not happen. This type of reasoning and imagining, which I will call *subtractive*, relies on the capacity to entertain mental representations that do not literally fit the world, i.e. imagine counterfactually. Subtractive reasoning is more complex than counterfactual imagination. While children are able to imagine many situations already around 18–24 months, the capacity to subtract from the current representation of the world one event e (or action) and to derive how this event not taking place would have affected reality (as it is currently represented by the child) is a skill that children acquire between 2 and 6 years. A more detailed discussion of the time span required for the mastering of this ability will be introduced in chapter 2.5.

An example of the distinction between counterfactual imagination and subtractive reasoning is the following: I buy a banana at a shop and I look at it. I can *imagine counterfactually*, i.e. against my knowledge about reality, that the object in front of me is a phone. This act alters reality and replaces some actual fact or property – i.e. the object is a banana – with some other imagined fact – i.e. the object is a phone. I can also *subtract* from my representation of the world the fact that I bought the banana, and try to reconstruct how my representation of the world would have looked like if I didn't buy it.

Notice that subtractive reasoning is essentially directed towards reality, in the sense that one imagines what would have happened if e didn't take place to compare this with how things stand actually. Quite differently, in the case of counterfactual imagination, the contrast with reality does not necessarily appear among the aims. In this sense, counterfactual imagination is self-sustained and motivated. Furthermore, notice that in order to imagine counterfactually the child doesn't need to fit her imagination within a certain given context, but the child can direct her imagination towards whatever is easier to represent. On the contrary, in order to engage in subtractive reasoning, the child needs to ensure that her imagined alternation fits in with her current representation of the world. She has to make sure that what she imagines not taking place, i.e. e, really does not happens in her imagining, and that the situation in which e does not happen is similar enough to the currently experienced one.

2.4.2.2 Back to the comparison

According to Nichols and Stich's proposal, imagination – which, for the absence of positive information against it, I assume to be conflated with pretence understanding – is essentially counterfactual imagination. Given a certain hypothesis, our inferential mechanisms filter what is consistent with it and construct a possible world[10]. This counterfactual imagination doesn't quite suffice to provide counterfactual reasoning or proper counterfactual statements of the sort "if φ were the case, then $\psi_1, \dots, \psi_n$ would also be the case", but it is involved in producing them.

According to Harris, counterfactual reasoning is essentially different from the imagination involved in pretending, since counterfactuals require a contrast between reality and an alternative (unreal, possible) situation.

[10] As previously mentioned, possible worlds in Nichols and Stich [85] do not correspond to possible world in the possible worlds semantics and modal logic, in that for Nichols and Stich possible world are allowed to be incomplete or inconsistent set of representation tokens.

I have the impression that Harris is more concerned with what I call subtractive reasoning, i.e. the ability to reason about what would have happened if something didn't take place, while Nichols and Stich seem to be more concerned with the ability to imagine something that differs from reality, i.e. counterfactual imagination. It is also reasonable to assume that the first ability is an essential element of pretend play, while the second is a distinguished ability that is not necessarily involved in pretending. To confirm these intuitions, however, it becomes even more urgent than before to dwell upon the literature in cognitive and developmental psychology concerning pretending and counterfactual reasoning. This will be done in the next section.

2.5 Developmental research on counterfactual reasoning

Various studies have suggested that young children are somehow able to engage in subtractive reasoning – i.e. to reason about how reality would have looked like if something didn't take place– quite early, but that around the age of 4 their performance is rather poor.

For instance, it is well known that children are usually able to pass the False Belief Task around the age of 4 (Wimmer and Perner [124]). Counterfactuals variants of the FBT suggest that children's ability to reason subtractively[11] is a necessary prerequisite to succeed in the FBT. This strongly suggests that children start engaging in subtractive reasoning around the age of 4 (Peterson and D.M.Bowler [89],Riggs et al. [96],German and Nichols [36]). Therefore, many authors have hypothesised that children start engaging in subtractive or "counterfactual" reasoning around the age of 4.

The results concerning counterfactual reasoning in early childhood are however much more diverse than it might appear at first sight: various studies conducted by P. L. Harris and others suggest that already 2-years old children are able to reason counterfactually – i.e. contra-known-facts –, if subjunctive questions concern invented content (J.Hawkins et al. [60]), or the material is presented in a fantasy context, such as a distant plant (Dias and Harris [23, 24]), if the premises are introduced with a dramatic make-believe intonation (Dias and Harris [24]), or if children are instructed to use imagery (Dias and Harris [24],Richards and Sanderson [95])[12]. Further studies suggest that alternative (non-verbal) modes of response (Beck et al. [5]), and considerations about causality (Harris et al. [45]) are also involved in facilitating counterfactual reasoning in early childhood.

Some authors have radically undermined the results hitherto mentioned, by showing that a full-fledged "sense for the nearest possible world" in counterfactual tasks is achieved roughly around the age of 12. These studies (Perner et al. [88],Cristi-Vargas et al. [20]) explain the previously mentioned results about counterfactual reasoning as the effect of "simple conditional

[11] Following the terminological difference previously introduced, I will call the ability to reason about what would have happened if something didn't (or did) take place "subtractive reasoning", and I will use the term "counterfactual" imagination to refer to the ability to imagine that something were different.

[12] For a review, the reader may consult Harris [43, chap. 6] and Leevers and Harris [70].

reasoning". According to the authors, children would employ simple rules to reason about situations that deviate from reality, rather than identifying the closest possible situations s.t. a certain event – which happened in reality– doesn't take place. The latter form of reasoning is called proper counterfactual reasoning, appealing to Lewis' work on counterfactuals (Lewis [73]). They defined counterfactual reasoning on the basis of a "sense for the nearest possible world":

> What abilities do children need to demonstrate proper counterfactual reasoning? Lewis (1973) proposed that, unlike normal conditionals whose truth depends on the real world, the truth of counterfactual claims is determined by what is true in the nearest possible world (that which resembles the real world most closely). Consider the sentence "If Carol had taken her shoes off, the floor would have stayed clean." This is true only if, in a possible world in which Carol had taken her shoes off (but is otherwise as similar as logically possible to the actual world), the floor would have stayed clean. The counterfactual claim is false if there is some nearer possible world, where Carol removed her shoes but the floor was still dirty –if, for instance, before Carol entered the room with her dirty shoes, her nasty brother had already messed up the floor with his dirty shoes. Hence, in this nearest possible world, even though Carol takes off her shoes, the floor is still dirty. (Cristi-Vargas et al. [20, p. 377])

Let me briefly present one of the tasks analysed by Perner and colleagues, to display why such a surprising difference is obtained within their study.

The second experiment in Perner et al. [88] modified a counterfactual task used in Harris et al. [45]. In the latter study, children were shown a clean floor, then they saw Carol coming home without taking her dirty shoes off, and making the floor all dirty. The past subjective question asked by Harris and colleagues ("If Carol has taken her shoes off, would the floor be dirty or clean?") was answered correctely even by 75% of the 3-years old. Perner et al. [88] hypothesised that a simple conditional rule of the form "If people take their shoes off (and the floor is clean), the floor stays clean" was used by young children to produce the response reported in Harris et al. [45]. Hence, they added to the story a second puppet, "Max", who walks with his dirty shoes right after Carol did so. In the modified condition, the question asked ("What if Carol had taken her shoes off? Would the floor be clean or dirty?") would have lead to the answer "dirty" under the counterfactual reasoning strategy, and to the answer "clean" under the basic conditional strategy. Among 5 and 6-years old children only the 18% of answers to the past subjunctive question (modified story) were correct.

Notice that the type of reasoning sought under the label of "counterfactual" differs across the literature: while most of the research claiming that young children can engage in counterfactual reasoning refers to the ability to think about something that contrasts reality (and what is known about it) or to reason on how things might have been different, Perner et al. [88] and Cristi-Vargas et al. [20] seek a full-fledged reasoning about causes, effects and how the removal of a cause (but nothing else) would have affected reality. Even though both uses of counterfactual are motivated, I will let the term "counterfactual" reasoning refer to the "sense for the nearest possible world", following Perner et al. [88], Cristi-Vargas et al. [20]. To denote the general ability to imagine that reality could have been different and to reason about these imagined situations, I will use the term "counterfactual" imagination. To refer to the ability to reason about how a certain event affects properties,objects and other events, I will use the term subtractive reasoning. Finally, to denote the ability to reason about content that contrast the agent's knowledge, I will make use of the expression contrary-to-fact (conditional or reasoning). On the basis of this distinction, it is possible to claim that children become able to imagine counterfactually quite early (already around 18-24 months), as it would be expected from the data about the offspring of pretend play. The ability to reason subtractively and contrary-to-facts, instead, seems to be still quite poor around the age of 4, but it is improved in make-believe, other worlds, or stories' conditions. Finally, as Perner et al. [88], Cristi-Vargas et al. [20] showed, the ability to reason fully counterfactual is mature around the age of 12.

For interpretative charity, I would like to claim that "counterfactual" reasoning in Perner et al. [88], Cristi-Vargas et al. [20] is not what Harris, Nichols and Stich are after. In order to establish this, however, a detailed analysis of why "counterfactual" reasoning is more difficult than subtractive conditionals and counterfactual imagination will be proposed in chapter 3.4 and in chapter 4.4.

2.5.1 Counterfactual reasoning in autism

Since pretend play is significantly different within the *ASD* if compared with neurotypical children (matched both by verbal and chronological age), it is interesting to focus briefly on the performance of children with autism in subtractive reasoning tasks. Let me start off from the counterfactual version of the FBT. Numerous studies have shown that children with autism fail the FBT long after their mentally-matched controls pass it (Baron-Cohem et al. [2]). On this basis, it has been claimed that children with autism's theory of mind module is impaired, and that they have difficulties in understanding other people's intentional states. The counterfactual version of the FBT displays that while 76% of the neurotypical children who pass the counterfactual version will also succeed in the (first-order) FBT, 44% of the autistic children who succeed in the subtractive task also succeed in the FBT. On contracts, both neurotypical and autistic children failure on the subtractive task is highly correlated with failures in the FBT (see Riggs et al. [96] and Peterson and D.M.Bowler [89]). This suggests that while subtractive reasoning is a necessary prerequisite to pass the FBT, there is more going on that represents an obstacle for children with autism.

Various studies (Dias and Harris [24],Scott et al. [104],Scott and Baron-Cohen [103]) have displayed that children with autism perform better than controls in subtractive and contrary-to-facts tasks[13]. If compared with verbal age matched controls, who are facilitated in abstract

[13] My re-labelling of the "counterfactual tasks", with the aim of avoiding confusion between the various terms that

subtractive and contrary-to-facts tasks by props to adopt a pretence or make-believe stance, children with autism do not show improvement under this condition. This shouldn't come at surprise, for autism is generally characterised in terms of deficits or impairments in pretend play (more details about *ASD* and pretence will be provided in section 2.6). It has been noted in Leevers and Harris [70] that children with autism are facilitated in subtractive task by a "yes" bias, which could explain why autistic children perform better than controls on the task.

In summary, it seems warranted to claim that children with autism do not display difficulties or impairments in subtractive and contrary-to-fact reasoning.

2.6 Pretence in children with autism

The Autism Spectrum Disorder is, according to the Diagnostic and Statistical Manual of Mental Disorders (4th edition) [26], a developmental disorder characterised by deficits in social interaction and communication, and by restrictive, stereotyped and repetitive behaviours and narrow interest. The *ASD* is sometimes distinguished from Asperger's syndrome, even though Asperger's syndrome is probably the mild end of the autism spectrum. I will, in any case, specify whether studies have been conducted on the former or on the latter whenever necessary. Autism is furthermore usually diagnosed, among other things, on the basis of the absence of pretend behaviour by 18 months. Since the lack of pretend play is one of the first signals of autism, a huge attention has been drawn to the study of pretence in autism. An overview of the most important families of theories about autism can be found in van Lambalgen and Stenning [117, pp. 241-46]. For reasons of space, I will here focus only on the most relevant hypotheses about *ASD* that have been explicitly related to pretence.

On the basis of what I have been saying so far, it should be clear that neurotypical children start pretending quite early, and seem to do so without needing any extrinsic motivation. They furthermore seem to enjoy the activity. Pretend play is often elicited by caretakers or playmates, who guide the unfolding of the make-believe episode. Lillard et al. [75] reports that "socio-dramatic pretending with peers emerges around 4 years of age, or earlier in the context of a more proficient partner such as an older sibling or the mother". I will call any instance of pretend play which involves more than one agent "social", in order to distinguish it from "individual" pretence. With the latter term I will denote episodes where a child is playing alone. I will further distinguish pretence understanding –i.e. the ability to identify which effects follow from pretence actions and transformations – from pretence production. The latter resides in the ability of the child to carry out some action that are appropriate with regards to the pretence stipulation(s).

Providing a unique and general definition of pretend play is almost impossible[14], but one

---

I have distinguished.

[14] A definition of play, in general, has been proposed in Burghardt [12]. After Piaget's work [91], the definition of pretend play has been extended to include the tree types of pretence defined in Leslie [71], i.e. object substitution, attribution of absent properties, imagination of absent entities. Furthermore, following Piaget [91], pretend play is

could say that pretend play involves the imagination of an alteration from reality, and a prop is used to represent another, non-present, imaginary target or a modified version of the prop itself (see Baron-Cohen [3]).

Children's varieties of play are generally classified as *sensi-motor*, *functional* and *symbolic*. The term sensi-motor play refers to the engagement in movements beginning with early reflexes and moving towards more intentional actions. With the expression functional play we mean the play where a toy is used appropriately, such as a spoon is used to feed a doll (see Bigham [8],Jarrold and Conn [57]). Finally, symbolic play is generally understood as the play where the child uses an object in a non-literal way, such as a brick is used as soap to clean a doll.

Pretend play, furthermore, can be deconstructed in the following three categories: object substitution, attribution of absent properties and imagining absent objects or entities. Let us try to carve out a general definition of these three kinds of pretend play, whose content may be compared with the formal definitions provided in chapter 4.2. These categories of pretend play have been famously proposed in Leslie [71, p. 414] and have since then become a standard distinction in the literature (see for instance Jarrold and Conn [57, p. 309]). Let me provide a definition of them, and an example.

An instance of imagination is considered a token of *object substitution* if one object is made stand in for another, different object (cfr Leslie [71]). The prime example of this type of play is using a banana as if it was a phone. Formally, object substitution means that a certain property of an object is replaced by another property. For example, the child's actions of grabbing a banana, holding it nearby her hear and uttering "Hey Daddy, how are you doing?" are to be understood as the child imagining that the thing they are holding is not anymore a banana but a phone.

Let's now focus on the case of *attribution of absent properties*. According to Leslie [71], when a pretend property is attributed to an object or a situation the play carried out is an instance of attribution of absent properties. Formally, I take this definition to mean that an act of imagination is an instance of *attribution of absent properties* if the agent's current representation of the world is augmented by some additional information, which was not true in the actual representation of the world. This means that while in the case of object substitution a property of an object is replaced with a conflicting property, attribution of absent properties refers to the addition of a certain property. For instance, suppose that a child sees a teapot. The child hence imagines that the teapot is hot. Notice that in this case it is not necessary that the child disregards a known property of the world (e.g. if she knows that the teapot is cold). Even though it is true that there could be some property of the world conflicting with the teapot being hot, it may also be that there is no information whatsoever concerning the teapot, or no information whatsoever concerning the teapot being cold[15].

often characterised as a symbolic play, even though there is disagreement as to whether the label is appropriate (see Jarrold and Conn [57, p. 309]).

[15]Recall that I haven't so far dwelt on the details of how the context dependent information is modified from reality to imagination. The conflicting information is an issue that will be extensively discussed in the sections concerning

Finally, we say that a child is *imagining an absent object or entity* if she introduces an object that is not in the current situation (and possibly some properties of it). For instance, suppose the child is sitting in the kitchen. She starts acting as if she was feeding herself with an spoon that is not there. This action introduces a new entity in her representation. In most cases – but not in full generality – it seems that imagining absent objects doesn't require to replace a known property of the world.

2.6.1 An overview of the experimental data

Pretence comprehension: various studies conducted in the last 20 years on pretence comprehension (e.g. Kavanaugh and Harris [62], Jarrold et al. [58], and Jarrold [56]) strongly suggested that children with autism are able to understand pretend transformations and actions. On the basis of these results, Harris and colleagues hypothesised that children with autism have a *performance* deficit affecting pretence production, while the competence for pretence is intact.

The results in Bigham [8] revert this impression, claiming that "while children with autism perform as well as controls when interpreting functional play, the comprehension of object substitution pretence and pretend gestures is somewhat problematic"(Bigham [8]). The surprising contrast between the results in Bigham [8] relative to Kavanaugh and Harris [62] are due to the following differences in experimental settings: the studies previously mentioned ([62, 58, 56, 59]) focused on the attribution of absent properties and imagination of absent entities. Bigham [8] also tested object substitution and pretence gestures, discovering that children with autism encounter difficulties in some types of pretend play. One hypothesis advanced by Bigham is that when target and prop share few or no characteristic features children with autism struggle in establishing a representational connection between them.

To the best of my knowledge, a worrisome paucity of studies testing whether autistic children understand the difference between pretence and reality have been conducted. However, there are some works on the reality-appearance distinction for both neuro-typical (see Flavell et al. [29]) and autistic children (see Baron-Cohen [4]). Below the age of 4, neuro-typical children score quite poorly on questions as to whether a rock-like sponge (which children could touch and squeeze) is really a rock or a sponge. 4-years old score definitely above chance, and 5-years old perform almost perfectly in answering the task (see Flavell et al. [29]). Furthermore, in Flavell et al. [29], phenomenist errors seem predominant when the reality versus appearance of the object properties were in question. Baron-Cohen [4] used a similar paradigm to the one employed by Flavell et al. [29] to test how children with autism perform on the reality-appearance task. While 81.3% of the normally developing group's answers were marked as correct, children with autism produced 35.3% correct answers. The author further reports that

> [i]t is also of interest that in those tasks that included plastic food, the autistic children alone persisted in trying to eat the object long after discovering its plastic

circumstances and subjunctive conditionals.

> quality. Indeed, so clear was this perseverative behaviour that the experimenter could only terminate it by taking the plastic object out of their mouths. In contrast, the normal and mentally handicapped subjects indicated they perceived the Experiment as some kind of joke or trick, showing laughter at the fake food and making comments like “It’s pretend chocolate!” (Baron-Cohen [4, p. 595])

On these basis, one could hypothesise that children with autism are more likely to score low on reality-pretence tasks than normally developing children. The results in Bigham [8] suggested that children with autism’s comprehension of symbolic play is impaired relative to controls, and that autistic children are more likely to produce reality errors when interpreting another person’s pretence behaviour. However, since it appears that little attention has been drawn to this topic, and with the aim of clarifying the role played by the reality-pretence distinction in pretend play understanding, in appendix B I will sketch a modification of Kavanaugh and Harris’ [62] paradigm that should provide a more definite answer.

In summary, it seems that children with autism are unimpaired in pretence understanding as far as the prop and the target share relevant characteristic features. However, in the case of object substitution and function play using a prop with a defined function different from the one it should stand in for, children with autism display some impairment in pretence understanding.

Pretence production: children with autism, contrary to a widespread platitude, are at times able to engage in pretend behaviours and carry out pretend actions. If compared with normally developing children, they display less spontaneous pretend play. Furthermore, they appear to spend less time pretending, preferring sensori-motor play and manipulations. The pretend behaviour is often described as more stereotypical, repetitive (Jarrold [56], Jarrold and Conn [57]). The latter report seems to fit with the hypothesis that autism is correlated with impairments in imagination and creativity (Craig and Baron-Cohen [19])[16]. When elicited, however, children with autism’s pretend behaviour improves, and the time spend engaging in pretence as well. The experiments carried out by Jarrols et al. [59] display that children with autism are impaired relative to controls in functional play – e.g. the use of a functionally appropriate prop to enact a pretend action – *contra* Leslie’s hypothesis that only symbolic play should be deficient. Furthermore, Jarrols et al. [59] shows that when instructed to enact a certain pretend transformation or action, children with autism are unimpaired relative to verbally matched controls. This suggests that autistic subjects do have an underlying ability to produce pretend play that is suppressed in spontaneous play. Symmetrically, as we have seen, Kavanaugh and Harris [62] supports the hypothesis that the ability to understand pretend actions is intact.

[16] As a methodological point, I would like to notice that impairments in imagination and creativity are often ascribed to children with autism *on the basis of impairments in pretence* (see for instance Craig and Baron-Cohen [19]). I suggest that this is problematic, and because of this conflation it is currently unclear whether subject with autism actually are impaired in their ability to imagine, or this is a difficulty in pretence understanding and production.

2.6.2 Theoretical hypotheses on autistic pretence

The *ToMM deficit-hypothesis* claims that children with autism display difficulties in pretending because of a meta-representational deficit (more precisely, impairments in the decoupling of primary representations). This hypothesis would also suggest that children with autism perform poorly on FBTs on the basis of the same impairment in decoupling primary representations. Furthermore, the ability to understand pretend transformation and actions displayed by children with autism is hardly explained by the ToMM deficit hypothesis. A discussion of the ToMM deficit hypothesis in relation with pretend play can be found in Nichols and Stich [85, pp. 47-57], Lillard [76],Scott and Baron-Cohen [103],Carruthers and Smith [15].

Following the *ToMM deficit-hypothesis* (Leslie [71, 72]) one would expect children with autism to struggle in symbolic play such as object substitution, while functional play should be unimpaired. As previously mentioned, the work of Jarrols et al. [59] challenges this hypothesis, by showing that in spontaneous conditions both symbolic and pretend play appear less frequently, and for less time if compared to verbally matched controls.

An alternative explanation of children with autism difficulties with pretence production (especially in spontaneous conditions) suggests that people within the *ASD* have an intact ability to engage in pretend play production and understanding, but they suffer from a *generativity deficit* (Jarrols et al. [59]). This roughly means that children with autism face more difficulties in generating ideas for pretence and pretend acts. On the basis of the possibility to overcome these problems through cues, Jarrols et al. [59] hypothesise that people with autism impaired production of pretend play might be due to a deficit in generating novel ideas *per se* or to a difficulty in generating strategies to retrieve ideas.

I will call *social-affective theory* a family of hypotheses and explanations about autism that focus on the different experience of social and affective interaction that children with autism seem to enjoy compared to normally developing children.

Hobson [49] proposed that the affective foundations of interpersonal communication are abnormal in autism. Through gaze and joined attentional activity, children start understanding themselves as social beings, separated by others. Hobson proposed that the valuation of these activities is different for people with autism. Many studies have observed that children with autism seem not to experience inter-subjectivity as rewarding or pleasing, and hence the development of skills that steam from inter-subjective relations are less advanced or abnormal. Even though Hobson [49] derives the cognitive symptoms of autism from the affective foundation impairment, it is possible to see the difference in social interaction and in pleasure in joint attentional activity as a co-cause of the conditions typical of *ASD*.

The *executive function deficit* (ED) is one of the general theories about *ASD* that is most often mentioned to explain children with autism performance in pretend play. According to Russell [99], the severe perseveration exhibited by people with autism and the difficulty in switching tasks is explained by a deficit in executive function, which is supposedly employed to maintain a goal. The

ED theory is deeply intertwined with neuro-psychological studies of patients both with autism and with frontal cortex damage. With regards to pretence, it is often hypothesised that autistic children pretence can be explained through an impaired response inhibition (Harris [42],Bigham [9]). Harris [42] hypothesised that the ability to impose internal executive control to contexts not supporting their plans is impaired. One can distinguish the ability to plan certain actions as external and internal: in the former, actions are guided by standard or habitual schemata evoked by the context, while internal executive control requires the devising of a plan tailored to the current task. Harris suggests that autistic children have difficulties in overriding external or habitual control spontaneously. This hypothesis is formulated with the idea of explaining the deficits in spontaneous pretend play, impairments in problem-solving and the difficulties in understanding mental states that are directed at a hypothetical situation that runs counter to current, known reality.

> [A]utistic children find it difficult to regulate their behaviour and their expectations in terms of a hypothetical context that runs counter to the current, immediately available context. Instead, they tend to fall back on habitual schemata prompted by the immediate situation. Harris [42, p. 244]

Prior and Hoffmann [93] and Hughes and Russell [51] (discussed in Harris [42] and van Lambalgen and Stenning [117, chap. 9]) provide evidence for children with autism's capacity for planning. However, the results in Bigham [8] and Bigham [9] seem to undermine the hypothesis that children with autism competence to pretend is marked by performance deficits, and that their difficulties are due to executive dysfunctions. The position I will try to defend within the present work is relevantly related to Harris' proposal [42]: as Harris does, I will relate autistic children's performance in pretend play with planning abilities and executive functions. In fact, Harris hypothesises that the problematic task for people with autism resides in over-riding external control spontaneously on the basis of internally devised plans, which do not match the external contexts. Contrary to Harris [42], I will try to explain pretend play within normally developing and autistic children on the basis of a dissociation between closed world reasoning about rules and about facts. Van Lambalgen and Stenning [117], Pijnacker et al. [92], Smid [107] and van Lambalgen and Smid [116] have found support for the hypothesis that people with autism have difficulties with closed world reasoning about facts, while they appear to have intact reasoning abilities about closed world reasoning about rules. Let me call this the "dissociation hypothesis". Before giving a taste of the distinction, and its motivation, let me say something about the reason for trying to apply this distinction to pretend play. As we have seen, all theories about *ASD* suffer from some objection and empirical counter-example. One might view this as a good reason to understand these different theories as complementary – even if addressing diverse levels of explanation – rather than competing. In any case, the aim of the present work is to display that the dissociation hypothesis is able to explain certain features of pretend play performance, both for neuro-typical and autistic children, without adding extra or *ad hoc assumptions*. Sufficient motivation has already been collected in favour of the dissociation hypothesis, and its application to pretence is an interesting generalisation of the explanation.

2.6.3 Closed world reasoning: the dissociation hypothesis

Van Lambalgen and Stenning [117] developed a logic for planning - diverging from classical logic with regards to semantics, syntax and consequence relation – that aims at modelling the way human beings reason about goals and plans, and hence the reason why humans fail to think according to classical logic. The details of this logic are sketched in appendix A, in order to facilitate the readers unfamiliar with the framework, and with the aim of keeping the content of this work as self-contained as possible (even though the curious reader is very much invited to have a look to [115] and [117]).

In many every-day life situations such as planning a certain action, interpreting someone's discourse, reading a train schedule, *et cetera*, our reasoning patters appear to be defeasible: a plan for instance is not provably true according to classical logic rules, i.e. true in every model of the premises. A good plan is rather true in the most probable situation. Furthermore, on the basis of people's performance on the Wason task, van Lambalgen and Stenning [117] suggest that rules are generally conceived as exception-tolerant. For instance, I may believe that the rule "If I am thirsty, then I drink" is true, even if in a situation where I know that the water available is poisoned or toxic I may be thirsty and not drink. Instead of thinking that the rule "If I am thirsty, then I drink" is falsified by this situation, many subjects claim that the poisoned water is an exception to the rule.

The term *closed world reasoning* refers to reasoning patters within which all information not explicitly stated or forced by the information stated is assumed to be false. Going back to the train schedule, if I am at Amsterdam's airport and the train schedule reports that there are no connections to Amsterdam after 3 a.m., I will (defeasibly) conclude that I cannot reach Amsterdam after 3 a.m. from Schipol's airport. However, I may later discover that there are also bus connections to Amsterdam, and that there is a bus leaving from Schipol at 4 a.m., contradicting my previous conclusion. Firstly, notice that the act of inferring consequences from the information available is non-monotonic: I can retract my previous conclusions and discover that on the light of new information they are not true anymore. Moreover, notice that I seem to be totally allowed to infer that there are no connections from Schipol to Amsterdam after 3 a.m. if I only know about the train schedule and nothing else. This can be seen as an example of closed world reasoning: I assume that everything not written on the train schedule is false. Later on, I do the same with the buses: I do not consider the possibility that the bus companies failed. I implicitly assume that if this were the case, the information would be reported somewhere nearby the buses' schedule.

Closed world reasoning can take two forms: one is reasoning about exceptions or abnormalities, also called closed world reasoning about facts (CWRf), the other concerns the rules that we "own" about a certain situation or object, or closed world reasoning about rules (CRWr).

In a simplified manner, one could understand CWRf as saying

If no exception is known, then assume no exception applies.

Continuing the train example, CWRf can be instantiated by the following reasoning pattern: when I look at the train schedule I implicitly reason that $(*)$ "if there are no trains to Amsterdam after 3 a.m., I cannot get home tonight". The form of this rule actually hides a lot of exceptions: if there are buses to Amsterdam after 3 a.m., or if I don't take a taxi, or if I cannot fly $\ldots$, then I cannot get home tonight. When I discover that there are buses to Amsterdam, I am dealing with an exception to the rule $(*)$. We engage with this type of exceptions most of the time, and we are generally quite at exception-handling. People with autism, however, seem to have more difficulties in integrating exceptions to a rule (see van Lambalgen and Stenning [117],van Lambalgen and Smid [116],Smid [107] and van Lambalgen and Smid [116]).

CWRr relies on a similar intuition to the one grounding CWRf, but concerning rules. One could state a rough version of the intuition behind it as

If ϕ is the case and $\psi_1 \to \phi, \ldots, \psi_n \to \phi$ are all the rules about φ, then assume that one of ψ_i is true.

This means that when reasoning about why a certain fact φ is the case, if certain rules are available – rules which would explain φ – then one does not normally hypothesise that φ is the case because of some χ which unbeknown to us entails φ.

Consider the following case, taken from Kowalski [65]: I wake up and notice that the grass in the garden is wet. I know that the grass is wet if it rained or if the sprinkler was on. Hence, if I want to find an explanation for the garden being wet, I will try to discover whether it rained or the sprinkler was on. I do not, however, check whether a rascal watered my garden to trick me, or think about all the possible rules that could be about the grass being wet. In doing this, I assume that only the rules – so to say– I know of or I have are valid, and everything else is false.

CWRf and CWRr collaborate in allowing humans to plan and go around in their daily lives, so they are usually intertwined. The experiments in Smid [107] on autistic performance on inferences, such *Modus Ponendus Ponens* (MP), *Modus Tollendo Tollens* (MT), *affirmation of the consequent* (AC) and *denial of the antecedent* (DA), suggests that CWRr and CWRf are dissociated in autism: the former appears intact, while the latter seems to cause difficulties.

The rest of the work will attempt at displaying how the dissociation hypothesis can be applied to pretend play, and to display which features pretending can be explained by it.

CHAPTER 3

A MODEL OF PRETENCE

Imagining alternative situations constitutes an essential part of human mental activity since early childhood. Children seem to engage in this activity for its own sake, and they enjoy it. As it has been displayed in chapter 1 and in chapter 2, the ability to imagine has been widely tied to counterfactual reasoning, in its various forms. In this chapter, I will develop a theoretical model (or better, a story) of the development of imagination (i.e. counterfactual imagination), pretend play and subtractive reasoning in childhood. Furthermore, I will attempt to propose a related story concerning the difficulties of autistic subjects to engage in imagination and pretence.

Henceforth, and throughout the rest of the thesis, I am going to talk about 4-years old or older children. This idealisation is going to simplify the discussion, since I will be able to focus on a specific time-frame of children's development. Furthermore, between 3 and 4 years, neurotypical children clearly display the ability to inhibit their knowledge about the world, as it is displayed by the False Belief Task (Wimmer and Perner [124]) and its counterfactual versions (Peterson and Riggs [90],Peterson and D.M.Bowler [89]). Hence, it is instructive to focus on children around this age, in order to clearly display how children with autism differ in pretend play.

3.1 Introduction

In chapter 2, Nichols and Stich's theory of pretence has been exposed and confronted with the one proposed in Harris [43]. Within this chapter, I am going to propose a story of the connections between imagination, counterfactual reasoning and pretence. In unfolding this story, various insights from Nichols and Stich, and from Harris, among others, will be included and discussed.

The choice of telling a story on this matter, instead of starting off with a full fledged theory, is motivated by various reasons. One of them is that the theory we aim to develop will inevitably be completed only in conjunction with the logical formalisation proposed in chapter 4. For this reason, a story seems to be an appropriate pre-theoretical approximation of a theory that will assume its final shape only by being put in relationship with subsequent formal modelling.

A second reason to prefer a story is based on the awareness that certain parts of the theory I aim to propose require further empirical investigation. An instance of this general phenomenon can be found in appendix B, where I sketch an experimental procedure on non-verbal pretence understanding. The structure of the task I propose essentially relies on the work of Kavanaugh and Harris [62], with the crucial difference that I propose to include a condition which tests the reality *versus* pretence distinction. I hypothesise that this task, as presented in appendix B, should fit in well with the results obtained in Jarrold et al. [58]. However, this example clearly displays that an ideal completeness of the present writing would require numerous experimental work. Moreover, the close relationship between theory and practice – which I aim for – entails that the theoretical model of the phenomena chosen may or must be modified on the basis of empirical results. Therefore, I am convinced that the story proposed hereafter could aptly be re-discussed and confronted with future work in the field.

Within this chapter, I am going to unfold a theory of imagination, subtractive reasoning and pretence. I am going to start off by displaying, in section 3.2, which type of imagination I will be analysing. In sections 3.2 and 3.3, I am going to argue for a theoretical distinction between *imagining something* (φ), and the circumstances in which the φ is imagined. This won't mean that I defend a view according to which imagining something is context-independent. Rather, in section 3.3 I am going to hint at the manner in which the general knowledge owned by a child, her beliefs about the world, her personal preferences and creativity can guide the process of imagining something. Moreover, the theoretical distinction between *imagining something* (φ) and imagining a situation making true φ will allow me to display how imagination differs from full-fledged counterfactual reasoning and subtractive reasoning.

Within section 3.4, the focus of the reader will be gently shifted towards subjunctive conditionals. I am going to display how subtractive reasoning differs from full-fledged counterfactual reasoning. Furthermore, I will display why full-fledged counterfactual reasoning – i.e. the "sense for the nearest possible world" described in Perner et al. [88] – is more complex than subtractive reasoning based on simple conditional rules. This will constitute an evidence for the suggestion advanced in chapter 2, i.e., that the reasoning described by Nichols and Stich as the basis of pretend play cannot be full-fledged counterfactual reasoning nor subtractive reasoning.

The chapter will find its close with the unfolding of the theory of pretend play I advance. In section 4.5, I am going to display how the theory of imagination and subtractive reasoning, developed in the previous sections, relates to pretend play. I am going to argue that imagining, subtractive reasoning and pretending share certain reasoning processes. Section 4.6 will display how the theory of pretence proposed for neuro-typical children explains the pretence production and understanding in the case of children with autism. In doing this, I am going to make essential use of the *dissociation hypothesis*, introduced in chapter 3.6, in order to show that its application to the case of pretend play induces fruitful hypotheses. Moreover, I am going to hint at how the dissociation hypothesis explains many experimental data about pretending in children with autism, data which have been introduced in chapter 3.6.

3.2 Imagining alternatives

Let us now forget for a moment the theories of pretend play that have been introduced in the second chapter (i.e. chapter 2), to think anew how to explain the development of the ability of pretending, and its relation with imagination and counterfactual reasoning. As we proceed, some connections and differences will be drawn with the theories previously exposed.

Firstly, as noted – among others– by Vygotsky, imagination is essentially related with reality and experience. The human ability to imagine forms of social and political life alternative to the ones that are and have been experienced obviously involves elements of novelty with respects to and difference from historically or directly experienced situations. However, they could not be generated without elements of our actual experience[1]. Vygotsky furthermore argues that reality and imagination are intertwined in that imagination broadens experience, allowing people to conceptualise something that hasn't been directly experienced[2]. The third factor is the influence of emotions on imagining and *vice versa* of imagining on emotions. Finally, the imagining acts we engage in may become part of reality and alter it.

The second core feature of imagination, on which I want to focus, is the spontaneity of imagining alternative situations. The literal processing of a situation or an event, just as much as the processing of a manifold of utterances, constantly leaves the door open to the alteration of the actual representation of the world. In a important matter, the alteration of a representation, and the anticipation of future outcomes of a situation or of a conversation, relies on the human capacity to plan. Both planning and imagining crucially involve the ability to engage in offline simulations (see van Leeuwen [118],Addis et al. [1],Schacter et al. [102] and Frith [32]).

The first building block that I need to introduce is the capacity to modify a representation of the world. Intuitively, it is hard to differentiate the capacity of altering a small component of a representation of the world from its integration with the circumstances. When I am asked to imagine that the coffee I am drinking were beer, I almost simultaneously carry out two "actions": I switch the actual fact that the cup contains coffee with the fictional fact that it contains beer, and I imagine a circumstance similar enough to the one I am experiencing where I have some beer in front of me. Even though these two elements may appear to be hopelessly intertwined, it is essential for the present purposes to distinguish them. Indeed, chapter 4 will display that

[1] For instance, Vygotsky affirms that "everything the imagination creates is always based on elements taken from reality, from a persons's previous experience" [119, p. 11].

[2] A similar point can be found in Harris [43].

they are naturally represented by rather different logical forms. Therefore, I am going to use the expression “imagine alternatives” merely to refer to the former phenomenon, and to investigate the latter in the following section (chap. 3.3).

Example: I am in the MoL Room writing my thesis (or pretending to do so). One hour ago someone asked me if I wanted to join them for a beer. Now I am tired of writing my thesis, and I think about what else could I have done to avoid the feeling of frustration about my thesis.

The fact that I was offered an alternative one hour ago, opens up an alternative situation to the one that I am experiencing. Since the situation I am currently in is frustrating, I can imagine what would have happened if I accepted the beer offer one hour ago. But this option does not exhaust the possibilities of my imagination: I can also indulge in thinking what would things be like if I join my friends for a beer now. Furthermore, because there are more fantasies that are dreamt of by our minds than things in heaven and earth, I can also engage in all sorts of imagining activities such as wonder whether I would have been happier if I never started writing my thesis, or what would have changed if I had decided to study at home today.

As it can be seen from our example, the type of imagination that I am seeking is much simpler than the creative imagination Vygotsky talks about. The ability sought here resides rather in the human capacity to alter the currently entertained representation of reality under some fashion. As mentioned, I will call this counterfactual imagination.

3.2.1 A variety of imagined alternatives

I now want to make sure that the ability to imagine an alternation and to construct some minimal circumstances for it is sufficient – in theory – to model imagination as pretence. A complete argument that the theory here proposed suffices to model imagination and pretend play will only be possible at a later stage, i.e. chapter 4. However, as we proceed in the unfolding a theory, I will try to display why the proposed notion can be understood as performing a certain function. The definition of the three important types of pretend play introduced in Leslie [71], i.e. object substitution, attribution of absent properties and imagination entities, can be found in chapter 2.6.

Formally, the logical framework that I will be using represents closed world reasoning, i.e. the intuition that only the formulae entailed by the current information available to me are assumed to be true. This will mean that the child will represent the tea-pot as actually full only if she has some positive information suggesting that it is. This will mean that, within the model proposed, to imagine absent objects doesn’t *necessarily* require replacing a representation: before the start of the pretence episode the child infers that there is no tea on the basis of absence of information about it. However, the fact that “the teapot is empty” is not added to the mental representation, and hence when the pretend play starts it is possible to simply augment the representation by adding positive information about the teapot being full.

The ability to entertain an alternative representation of reality – i.e. of modifying our current representation of the world in some manner – is general and wide enough as to be in principle able to cover the three cases of object substitution, attribution of absent properties and imagination of absent entities. As it can be noted by the intuitive definition proposed in chapter 2.6, all these activities involve the ability to imagine that something is different from how it stands.

3.2.2 Imagining the impossible

The question of whether human beings are able to imagine, or better *positively conceive*, impossibilities is the matter of furious debates. Within the present writing, I have been unable to devote much space to this issue, since I gave priority to a general account of how imagination in pretence and subjunctive reasoning works. Nonetheless, the reader of this work might consider many instances of pretend play and imagination, which I analyse, as *impossible*, either logically impossible or metaphysically impossible.

As a methodological choice, I decided to avoid imposing boundaries to what children seem to able to imagine in terms of metaphysical and logical necessity or impossibility. I am convinced that the discussion on the relationship between imagination and modality hinted at in the introduction (chap. 1) is possible only when a plausible theory of the phenomena under consideration has been proposed. Therefore, whether the impossibilities children *appear* to imagine are in fact impossibilities or not exceeds the scope of the present work: I merely want to point out that children learn to imagine, to reason counterfactually, and to pretend by imagining things that seem impossible.

For the claim that many examples of pretend play could be considered metaphysically impossible, consider an essentialist, who believes that being a banana is an intrinsic and essential property of the objects that we call bananas. She would claim that pretending that the banana is a phone exceeds metaphysical possibility. Similarly, according to Saul Kripke, even though identity statements may give the impression of being contingent, they are necessary. Therefore, the fact that I can think that the table in front of me could have been made of ice instead of wood and my conviction that I can *imagine* this doesn't entail that I am imagining *this* table. Kripke suggests that this is an illusion: I am imagining a table with the very same position and with the same phenomenal appearance of this one, except that it was in fact made of ice, i.e. a doppelgänger of the object[3]. Hence, according to Kripke, we may end up thinking that certain identities are contingent because we imagine something else with its phenomenal qualities. As plausible as this may sound, it is unclear whether this theory applies to the representations entertained by a human being imagining herself. As it has been argued in Berto and Schoonen [7], on the basis of the phenomenological access that humans have to themselves it sounds implausible to claim that I am misidentifying myself on the basis of phenomenal qualities and appearances.

[3]See Kripke [67, p.160-161]: "[O]ne could have the illusion of contingency in thinking that this table might have been made of ice. We might think one could imagine it, but if we try, we can see on reflection that what we are really imagining is just there being another lectern in this very position here which was in fact made of ice. The fact that we may identify this lectern by being the object we see and touch in such and such position is something else".

Taking at face value the numerous reports and direct experiences of children pretending to be a dead cat, an animal, a train even (to mention but one example see [85, p. 20]), it seems that pretending and positive conceiving[4] can exceed the boundaries of the possible, in this *metaphysical* sense. Whether pretending to be a cat (or imagining to be a cat) means that I imagine myself having the DNA of a cat, or that I imagine the some typical cat-like acts and properties is not determined by the present opera, and it is open for debate. In any case, it seems more fruitful to let children imagine whatever they can imagine, so as to formulate a comprehensive theory of imagination, before deciding what can or cannot be imagined.

So far I have focused on metaphysical examples, but impossibilities can be further understood in a logical fashion: imagining a logical impossibility could correspond to imagining a flat contradiction, or imagining that every model of Peano Arithmetic is standard. Notice that also the possible worlds used by Nichols and Stich [85] are in the logical sense impossible: the authors aim at referring to descriptions which may contain unobvious contradictions such as "there is a greatest prime number". The formalism that I will use in chapter 4 does not allow imagination of obvious contradictions, because negative information is viewed as absence of positive information. If one wanted to formalise the imagination (or better, positive conceiving) of logically contradictory sentences, the formal approach that I will develop in chapter 4 could be easily extended as to include contradictory propositions.

I will briefly sketch how the formalisation I propose can be augmented to model imagination of impossibilities. The details heavily rely on the work of Robert Stärk (Stärk [108],Jäger and Stärk [55]). The basic idea behind Stärk semantics is to extend the classical three-valued semantics used in Logic Programming by adding a fourth value, in order to obtain a truth value that stands for contradictory. The set of truth values, following Fitting [28], is $Four : \{0,1\} \to \{0,1\}$, where $\langle x, y\rangle \in Four$ means that x stands for the degree of truth and y stands for the degree of falsity. Hence, the pair $\langle 1,0\rangle$ corresponds to truth t in Kleene's semantics, $\langle 0,1\rangle$ corresponds to Kleene's falsity f, $\langle 0,0\rangle$ stands for Kleene's undetermined, and the pair $\langle 1,1\rangle$ stands for contradictory. A partial ordering is defined on $Four$ by

$$\langle x, y\rangle \leq \langle x', y'\rangle \Longleftrightarrow x \leq x' \text{ and } y \leq y'$$

The structure $(Four, \leq)$ is a lattice (displayed in fig. 3.1). Stärk proves that the four valued semantics defined the three-valued consequences and four-valued consequences of the completion of a program are the same[5].

For the present aims, this means that the formalism here proposed could easily model the imagination of contradictory propositions, without too many changes in the logic. A more detailed exploration of the implications of this extension, however, are left for further work.

[4] Recall that the imagination I describe corresponds to positive conceivability, for imagining that φ always involves a situation verifying φ.

[5] For any formula without equivalence. See Stärk [108].

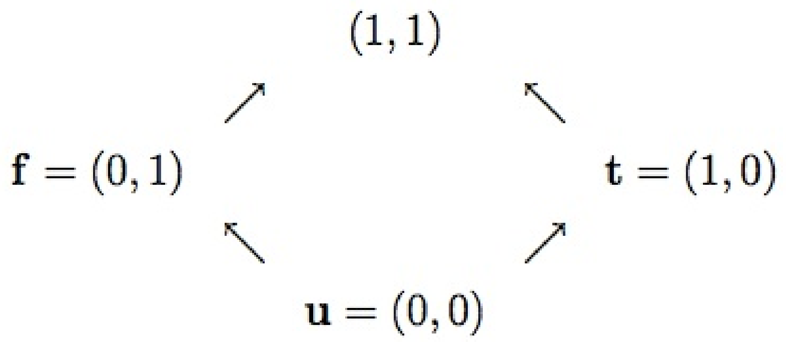


Figure 3.1: The lattice for Stärk four valued semantics for Logic Programming

As a concluding remark, it shall be noticed that even though within the present work a three valued semantics is preferred to the Stärk's four valued semantics, this doesn't mean that imagination of impossibilities is prohibited. I previously mentioned that many instances of pretend play can be understood as metaphysically impossible. Furthermore, as it was hinted in 3.4.2 (see also the appendix A.3.5), the minimal models and circumstances used are not maximally consistent sets of formulas. They are rather strongly incomplete, according to the idea behind Closed World Reasoning: all information that is not contained nor forced by minimal rules about causality and about typical scenarios are assumed to be false. This means that the minimal circumstances here used could be understood as impossible worlds of certain types (cfr. Berto [6]). Moreover, the imagination hitherto described is already *impossible* in the sense that it can violate both logical prescriptions – i.e. one can imagine that a situation s.t. neither φ nor $\neg\varphi$ is the case – and the most general (causal) rules about the world, as it will be displayed in chapter 4. As I claimed at the beginning of this section, the question as to whether this sense of "imagining" really violates any form of absolute, logical or metaphysical, necessity, is left undetermined. This is not to claim that the discussion as to whether we can truly imagine or positively conceive impossibilities is not worth pursuing, but rather that only once a general and comprehensive theory of how imagination works and how children learn to imagine has been formulated, the question can be sensibly addressed. Indeed, analysing how children's imagining can relate to the debate on imagination of impossibilities in metaphysics would constitute an interesting extension of this writing.

3.3 Circumstances

In the previous section, I assumed that the imagination of an alternative can (and should) be distinguished from the circumstances in which we imagine the alternative (and their relationship with the actual circumstances). I will now try to argue for this assumption, only once this aim will be achieved I will display how the imagination of the circumstances relates with the imagination of an alternative.

The claim that *imagining* φ, for a proposition φ, shall be distinguished from *imagining the circumstances for* φ shouldn't be understood as affirming that when we imagine φ we do not imagine some circumstances in which φ is true. Quite on the contrary, I am convinced that our imagining that φ always involves some circumstances. The distinction drawn is meant to disclose that our imagining that φ is not necessarily imagining that φ is the case in the closest possible circumstances to the actual ones. With the *closest possible circumstances* I refer to a representation of the world that is exactly as the one currently entertained, except that with regards to φ and what follows form it [6]. Furthermore, let the term *minimal circumstance* (or minimal model) denote an incomplete model of φ s.t. it only makes true φ and what follows from φ and the most general rules about the world that the agent owns[7]. I will suggest that imagining a minimal situation is more fundamental than imagining the counterfactual situation.

Looking at examples of imagination in early childhood it appears that kids have the ability to entertain a representation of φ with some minimal circumstances that make φ true, and to devise a preferred context for their fantasy that φ on the basis of personal preferences, subsequent developments *et cetera*. For example: a child playing with her puppets and imagining that they are getting married doesn't seem to imagine the closed possible situation where the two puppets involved are getting married – if there is anything such as the closest possible world where the two puppets are getting married. She often seem to start simply imagining that they are getting married, and to derive a description for the circumstances in which they are. In this sense, it seems that children can modify their representation of the world as little as possible as to accommodate the imagined fact that the puppets are getting married – e.g. they talk, are like humans *et cetera*–, just as much as they can entertain a highly partial representation, making true only that the puppets are getting married. In the latter case, children seem to construct a circumstance where it is likely that the puppets get married on the basis of their knowledge – however limited – of the world and of their creative preferences.

Leaving aside the further decisions that I can impose on my own imagination, the crucial realisation is the following: when we imagine that something φ is the case, we are able to imagine a variety of circumstances in which φ is true. The circumstances can be as close as possible to reality, but they may as well deviate from reality under various manners, or be highly incomplete. Therefore, it makes sense for a formal point of view to distinguish the imagination of the closest possible context to the actual one such that my supposition (or imagined fact) is true, and a minimal context that makes true my supposition. Even though acts of the imagination rely on reality to arise, imagination doesn't have to stick around reality too much.

[6] This is what Perner et al. [88] and Cristi-Vargas et al. [20] call counterfactual reasoning.

[7] A more detailed introduction to the ideas behind minimal models can be found in section 3.4.2 and the part of the appendix concerning minimal models in Logic Programming (sec. A.3.5).

If our imagination were always counterfactual in Perner et al. [88] sense[8], then we would have to constantly retract from our previous counterfactual suppositions to enlarge them with further hypotheses. Instead, it seems that we often just imagine something – e.g. that my shoes are red – without specifying much context of this imagining. It seems that we can just imagine some minimal circumstances in which the shoes are red, such as they are still mine, maybe they are at my feet and they are red.

Let's suppose that it is granted to us that our ability to imagine alternatives does not necessarily nor mostly come in the form of the imagination of the closest possible situation to the actual one. Then, it is clear that the counterfactual imagination displayed in early childhood is not of the counterfactual kind à la Cristi-Vargas et al. [20].

This doesn't entail that there are not important connections between subtractive and contrary-to-facts reasoning and imagination: the following sections are aimed at displaying that how I envisage this connection.

3.3.1 How to construct the circumstances

I have argued that imagining that something (φ) were different is not necessarily imagining that something is different in the closest possible situation. However, I have so far said very little about how the construction of some situation that makes φ true shall be characterised. This section attempts at compensating for this deficiency, by displaying which mental abilities and sort of knowledge may be involved in constructing some plausible circumstances.

The most prominent candidates that concur in constructing a plausible situation for a supposition ϕ are *creative imagination*, *causal knowledge* and *general knowledge*[9]. I will start from causal and general knowledge, for they seem easier to deal with.

Let's suppose that a child Jenny imagines that the sky is red. According to my theoretical distinction between imagining that φ and imagining the circumstances, the ability of imagining that φ is true (or false), while it is false (or true), can (and should) be differentiated from the ability of imagining in which circumstances φ is true. Hence, Jenny's imagining that the sky is

[8] That is, if we were always imagining a situation that is exactly as the actual world except with regards to an imagination premise when engaging in imagination and pretend play. "Counterfactual", as I claimed in chapter 2, refers to the definition of counterfactual reasoning or sense for the nearest possible world proposed in Cristi-Vargas et al. [20]. Perner, Rafetseder and Cristi-Vargas argue that "unlike normal conditionals whose truth depends on the real world, the truth of counterfactual claims is determined by what is true in the nearest possible world (that which resembles the real world most closely)". Even though, following Lewis [73], it is not necessarily the case that there is *one* nearest possible situation, I will grant this terminology to Perner et al. [88], Cristi-Vargas et al. [20]. I will hence use the terms subtractive and subjunctive reasoning as to refer to the ability to engage in reasoning processes about some situation that in intended to be different from the currently experienced reality.

[9] Notice, *en passant*, that the usage of causal and general knowledge within this context does not entail that the child is aware of the knowledge's clauses she is using, nor that she must have any specific knowledge about φ. Consider for instance a child pretending to be a flamingo. She may not know much about flamingos specifically, but that they are a type of birds. In this sense, I follow Doggett and Egan [25], in claiming that knowledge *guides* imagination (also the one in pretence) rather than *completely determining* it. In any case, it is true that imagine that "grimles grimlesise" is rather hard to do, if I don't have the faintest idea of what a "grimle" is, or to which context I could relate it to. This changes completely, obviously, if I make up the word grimle and I decide what this should refer to.

red does not automatically determine the context imagined by Jenny when she imagines that the sky is red. However, if Jenny knows that the sky can only be red because the sun is setting, then she may want to imagine that the sky is red in the circumstances where it is setting. In the case where the general knowledge, owned by, and the scripts, of typical behaviours and situations that Jenny knows, allow her to identify some minimal and unique conditions that would cause the sky to be red, then it is relatively easy to construct a context for Jenny's imagining that the sky is red. On the other hand, if there are many multiple conditions that may cause or entail that the sky is red, then Jenny should be able to pick one of these conditions. This choice may be guided by creative imagination and personal intentions – such as Jenny's deciding to break the convention that sky is red at sunset, and to imagine that it is midday and the sky is red –, but it could also be guided by the maxim of finding the minimal explanation that allows to explain as much as possible. In the latter case, it might be that Jenny was imagining that the sky is red and there is a spaceship floating in the sky, then the hypothesis that aliens are coming down to Earth may be a minimal explanation. For instance, Jenny may decide that if (in normal circumstances) alien bodies approach the Earth this changes some chemicals in the Troposphere. Then, the minimal circumstance would include that the sky is red, that there is a spaceship floating in the sky and aliens are approaching the Earth. Together with the above-mentioned rules, this would explain all the imagined evidence.

It might have already occurred to the reader that I conceptualise bits of the construction of imagined circumstances as similar to an abductive process, i.e. inference to the best explanation of some unexplained observation. The usage of abduction to model the construction of the context for an imagined event is driven by the analogy between unexplained effect and imagined event. In abductive reasoning, we reason from unexplained effect seeking for a relevant cause or set of causes that could explain what we observe. In setting the circumstances in imagination, we start from an imagined event and we want to find some relevant situation in which this imagined alternative may be true. In both cases, we reason backwards from a fact to an explanation of it. Furthermore, in both cases it is essential that the explanation found is relevant to the event, for otherwise nothing would be explained. Finally, the acknowledgement that abduction on its own cannot represent everything that goes on in imagination does not entail that it cannot be used to understand and formalise some similar phenomena within the field of imagination.

As far as creativity and personal decisions are concerned, it is clearly quite hard to explain why and how people decide to guide their mental life, not to mention their imagination. Furthermore, even though children can sometimes direct their imagination towards what they prefer or towards something intentionally different from reality, it should also be noticed that they are not always able to do so. When a child is scared by an imagined monster, she may be unable to control her emotions triggered by imagining. However, notice that children often avoid imagining and pretending certain things because they are too scary (see Nichols and Stich [85, p. 20]). This suggests that children can direct their imagination, even though the emotional effects of an imaginative act are not always under voluntary control.

I will introduce the formal counterpart of what has been discussed so far in chapter 4.3. However, it can be already anticipated that the theory I am proposing leaves open the possibility of intentional intervention on imagination. This does not entail that I assume that humans are always able to intervene on their imagination activities, but they sometimes manage to do so.

3.4 Subjunctive conditionals

The ability to think what would have been the case if an event e happened if it actually didn't, or didn't happen if it actually did, is what I call subtractive reasoning. The ability to engage in subtractive reasoning seems to be more complex than simple imagination. When I reason about what would have been if something were different, I do not merely imagine that something is different (e). I also need to understand how e being the case or not affects the rest of what I represent as the actual situation. Furthermore, I need to make sure that everything in my actual representation that entails or causes e is removed: in a big enough representation I could have that e is caused (or entailed) both by ψ_1 and by ψ_n. Hence, I cannot suppose that e is not the case without backtracking that both ψ_1 and ψ_n must be false as to now make e false.

Checking whether any ψ_i in my representation of the world may entail (or cause) e, however, might be much more involved and time-consuming that just imagining that e – with some minimal circumstances – is false.

I suggest that there is a sense in which children below age of 12 are somewhat able to engage in subtractive reasoning. Following an idea of Mackie [79], before developing a full-fledged ability to create the closest circumstance(s) to the actual ones, children reason on the basis of imagination, analogy and typical scripts. Hence, as recently argued by Perner et al. [88] and Cristi-Vargas et al. [20], children before schooling are not able to engage in full-fledged "counterfactual" reasoning. With the latter I refer to the imagining that φ was true if it is actually false and *vice versa*, altering the representation of reality as little as possible except with regards to φ. Rather, on the basis of simple rules and of analogy between situations, children are able to modify the actual representation to suppose that something is different. However, in doing this, the use of general rules, scripts and analogy does not put the child in the best conditions to always identify the closest possible situation.

In the parallel section in 4 (chapter 4.4), I will display why subtractive reasoning is more expensive and requiring than imagination and pretence. I will furthermore mention why counterfactual reasoning à la Perner et al. [88] is more complex than the subtractive reasoning displayed by 4 years-old. For the moment, I will restrict my attention to some results within the cognitive and psychological literature that provide some support for the hypothesis that "counterfactual" reasoning – insofar as it is understood as the "sense for the nearest possible world" – is harder than simple imagination. The following chapter, moreover, will be devoted to shedding light on the connection of imagination and pretend play. There, I will argue that counterfactual imagination and subtractive reasoning predates counterfactual reasoning because pretend play is cognitively less requiring.

3.4.1 Proper counterfactual reasoning is cognitively requiring

I approach the explanation of why counterfactual reasoning should be more demanding than subtractive reasoning, by referring back to some results introduced in section 2.5. Notice that in what follows I assume that counterfactual imagination is simpler than and the basis for both subtractive and counterfactual reasoning.

The studies conducted by Riggs et al. [96],Harris et al. [44] test children's ability to identify "how the world would be now had some earlier event not occurred"(see [96, p.75]). To correctly reason about this type of question, subjects must be able to suppress knowledge of the current reality, and to obtain a representation of a situation that is different but is meant to stand in for current reality. However, because it is implausible to assume that children aged 3 to 5 have a complete knowledge of causality, the results of Riggs et al. [96] and Harris et al. [44] can be accounted for by assuming that the children

1. are entertaining an alternative representation of the world,
2. are able to create this alternative representation of the world on the basis of reality, but
3. the creation of the alternative is guided by some simple conditional rule, imagination, scripts stored in episodic memory and by analogy.

Perner et al. [88] and Cristi-Vargas et al. [20] tested whether children are able to engage in a full-fledged "counterfactual" reasoning task, where "counterfactual" reasoning is defined as the "sense for the nearest possible world", i.e. reasoning about an alternative representation obtained from reality and compared with it, which differs from reality as little as possible. The above mentioned studies have devised some experimental conditions that allow to distinguish the "proper" "counterfactual" reasoning à la Cristi-Vargas et al. [20] – which could also be called the sense for the nearest possible world – from what they call basic conditional reasoning – which corresponds to what I have hinted at saying that the creation of an alternative is guided by imagination, simple rules and analogy.

> [...] unlike normal conditionals whose truth depends on the real world, the truth of counterfactual claims is determined by what is true in the nearest possible world (that which resembles the real world most closely). Consider the sentence "If Carol had taken her shoes off, the floor would have stayed clean." This is true only if, in a possible world in which Carol had taken her shoes off (but is otherwise as similar as logically possible to the actual world), the floor would have stayed clean. The counterfactual claim is false if there is some nearer possible world, where Carol

> removed her shoes but the floor was still dirty –if, for instance, before Carol entered the room with her dirty shoes, her nasty brother had already messed up the floor with his dirty shoes. Hence, in this nearest possible world, even though Carol takes off her shoes, the floor is still dirty. With "counterfactual" reasoning, we would thus conclude that the floor would have been dirty, a conclusion that basic conditional reasoning would not have produced (see [20, p. 377]).

The results of Perner et al. [88] and Cristi-Vargas et al. [20] display that children below the age of 6 mostly solve counterfactual tasks on the basis of simple conditional reasoning: they follow a simple rule about the world and they construct an alternative on the basis of this simple rule. Cases of co-causation – i.e. in circumstances where two events concur in causing a certain effect (but they may happen at different times) – already suffice to show that children use a simple conditional rule, according to the authors. For instance, if the experimenter(s) tell the following story about Simon and Julia, two siblings.

> Simon and his little sister Julia both have their own room and each likes really much candies. Their mum brings a candy in the kitchen, and she places it in the box on the top shelf (alternatively: in the box on the bottom shelf). Simon is tall enough to reach the top shelf, but he has his leg in a cast and could not kneel down to reach the bottom box. Julia can reach the bottom shelf, but she is not tall enough to reach the top shelf! If Julia found candy on the bottom shelf, she brought it into her room, and if Simon found candy on the top shelf, he brought it into his room.

After familiarisation with the settings and the characters, the participants were asked a memory and a future indicative question, respectively "where is the candy now?" , and "what will happen when Simon comes to look for a candy?", just after the mother brings a candy in the kitchen. Then, either the boy or the sister comes to look for a candy. The children are asked an other memory question, "Where is the candy now?", and a counterfactual question, "What if the little girl had come looking for the candy instead of the boy? Where would the candy be?". In the critical condition where the candy is placed on the top shelf (hence only Simon can reach it), and he comes first, subjunctive questions were particularly hard (6% correct). Even though Julia cannot reach the top shelf, children most commonly judged that the candy would have been brought to her room.

This example gives a flavor of the interpretation drawn by the authors: children reasoned on the basis of a simple conditional of the form "If character X comes first, then s/he takes the candy to his/her room".

Taken together, the studies of Perner et al. [88] and Cristi-Vargas et al. [20] provide some evidence that the sense for the closest possible world is not fully developed until age of 12. It is relevant to mention that in the story just mentioned, the solution of the task observed in young children was also sometimes present among adults: "[. . .] 6 people showed the [. . .] pattern we described for the small group of children in Experiment 2. When the object was counterfactually assumed to have been in a location where the actual character could not reach it, the participants assumed the character who could reach there would eventually come and fetch it." This suggests that there may be interpretative problems, i.e. problems in reasoning towards an interpretation, for both children and some adults. However, even the non problematic conditions for the same experiment seem to support Mackie's characterisation of subtractive reasoning. In Mackie [79], it is argued that our causal knowledge relies on the ability to imagine counterfactual situations: we do not reach causal knowledge simply from observing regular events, rather we engage in thought experiments in which we imagine what would have happened in a specific circumstance had something not happened. By comparing the alternative cases where the observed outcome did not occur, we identify the causal link.

Mackie's theory also explains why what we consider a cause for a similar (or the same) event changes on the basis of the circumstances: the causal conclusions we draw really much depend on the alternatives we are going to consider, which rely on the circumstances under consideration. This can be paraphrased by saying that we generally tend to consider something as a cause if it is not typical in the circumstances. If a fire breaks in a laboratory, we might consider a light cigarette being abandoned in the factory a cause, while with all probabilities we wouldn't consider the presence of oxygen as a cause (even though it is weakly necessary for the fire to break out). However, if we add that the laboratory has special precautions to avoid the presence of oxygen during experiments, we may quite promptly consider the presence of oxygen as a cause.

It is interesting to notice that the idea that causal links are drawn on the basis the circumstances and what is typical within them nicely fits with the experimental data according to which children are more prone to reason counterfactually when a relevant alternative is introduced. I will further develop on this idea in chapter 3.4.2.

Mackie proposes to distinguish two approaches to the creation of counterfactual alternatives and their comparison: a mature and an early strategy. The early strategy relies on imagination, scripts and analogy, while the the mature construction of alternatives and comparison is based on general inductive propositions concerning what would happen in a contrary to fact situation and deduce certain consequences. Therefore, following Mackie's theory, both proper "counterfactual" reasoning à la Cristi-Vargas et al. [20] and simple conditional reasoning – as it is characterised in Cristi-Vargas et al. [20] – deserve to be called subjunctive reasoning[10]. By adopting his account, it is possible to read Cristi-Vargas et al. [20] and Perner et al. [88] as a precisificaiton of Mackie's theory and of many studies of children's ability to reason counterfactually. Before the age of 12-13, humans are imperfect counterfactual reasoners, and they make use of imagination and analogy to construct the relevant alternatives. The basic conditional reasoning, on the other hand, relies on counterfactual imagination, which – as we have seen – largely predates subtractive reasoning.

[10] They actually deserve to be called "counterfactual", in a broad sense, but I used this term to denote the "sense for the nearest possible world" sought by Perner and colleagues.

3.4.2 Subjunctive reasoning is prompted by the introduction of (unrealised) alternatives

Various authors have noted that humans are more prone to reason counterfactually when the situation they are processing includes some relevant alternative. Byrne [13], Harris [43] and Harris et al. [45], among others, have suggested that people are generally more likely to engage in subtractive reasoning when one (or more) relevant alternative(s) to an (undesired) event is mentioned. In order to display how subtractive reasoning may originate – or to offer a story on why subtractive reasoning is facilitated by relevant alternatives – we will adopt the Jenny story from a counterfactual task present in Riggs et al. [96].

> Jenny is in the garden painting a picture. She considers leaving the picture on the table or bringing it inside. At the end, she leaves the picture on the table, and the wind blows it up into a tree.

The control question would be "where is the picture now?", while the test question would be "How could have Jenny avoided the painting being on the tree?".

The formalism that will be introduced in 4, on the basis of the definitions introduced in appendix A, is meant to represent the way humans generally construct minimal representations of both what they receive as sensory input, and of discourse interpretation. A minimal model of a program P, where a program is supposed to represent a certain representation of the world, is a model making true only the formulas in P and what follows from the axioms of causality. The program P may include scripts or rules about a certain scenario, in which case its minimal model $T_1(P)$ will also make true what is forced to happen by the scenario.

For the moment being, I will borrow the terminology of minimal models, to simply refer to the partial representations we entertain as the environment around us is processed. In the case of Jenny's story, this simply means that the hearer of the story constructs partial representations – which I will call minimal models – as the story is unfolded. The minimal models however, do not grow linearly. They may be contradicted or modified by further utterances or events (see [115, p.84]). Therefore, at the point the hearer is told that

> (P) Jenny is in the garden painting a picture. She considers leaving the picture on the table or bringing it inside.

the minimal model $T_1(P)$ will not contain nor that the picture is left outside neither that it is brought inside. However, at time t the hearer of the story will consider as equally possible next minimal models the ones where the painting is brought inside $T_2(P)$ or the one where it stays outside $T_3(P)$. The continuation of the story forces the hearer to accept the minimal model where Jenny's painting is outside. Furthermore, this model is later extended as to represent that the wind blows up the painting and it ends up on top of the tree.

When the hearer of this story is asked "how could have Jenny avoided the painting being on the tree?", according to the story so far proposed, she imagines an alternative to the events narrated within the same circumstances. However, the choice of the relevant alternatives is facilitated by the fact that at time *t* she considered plausible that the painting was brought inside. Hence, she can backtrack to that minimal model she entertained and check whether there the wind still causes the painting to end on the tree or to move at all. Notice, furthermore, that the logic I use is meant to capture what is typical and what's abnormal in a given circumstance: this entails that the logical approach here advanced already captures that untypical or abnormal events are more likely to be considered, in counterfactual reasoning settings, as the causes for undesired effects. A more detailed explanation of these phenomena in logical terms will be offered in chapter 4.4.

Children's ability to imagine counterfactual scenarios to the one they are literally processing – let that be the case of reality' processing or of discourse representation – can be represented on the basis of the temporal dimension of the interpretation of the world. Humans' representations get generated in time and they are modified (extended or retracted) on the basis of further information. At each step of the processing of the world, people are able to foresee some of the developments that their representations are most likely to undergo. Let us consider the example of discourse representation: if I tell you that "John fell. Max pushed him (back).", theoretically we can identify various steps of processing. When you only have been told that "John fell", you construct a minimal model which only contains the fact that John fell and minimal context dependent assumptions – e.g. that John is the only person so called that we both know *et cetera*. Even though there is only one minimal model of the discourse processed, there are some plausible minimal models that could follow, such as that John fell and hurt himself, that John fell because someone pushed him, that John fell because he stumbled upon some un-seen object, that John fell because he was having a fight with someone.

The later introduction of the second part of the sentence "Max pushed him" forces the modification of the minimal model at this later stage with the fact that John fell because Max pushed him. Notice, as it is explicitly displayed in van Lambalgen and Hamm [115], that we infer that Max pushing John causes John to fall on the basis of our general and causal knowledge. In this specific case, for instance, because we know that pushing someone generally leads to this someone falling or being thrown in a certain direction with a certain force and so on and so forth.

Finally, the addition of "back" in "Max pushed him back", would force the agent listening to me to retract from the minimal model in which John is peacefully doing things and Max aggressively pushes him, to a minimal model where John was having a fight with someone, Max in this case, fight that John himself started by being aggressive, for instance.

In our story, what happens is that my interlocutor goes back to the first minimal model that she computed of the utterance I am making to construct as the new minimal model the one – which she considered plausible before – where John had a fight with someone. This means that agents are generally able to access their precedent representations.

As it is known from the Smarties' task (see Perner et al. [87]), however, kids below the age of three have difficulties in reporting what they believed at a precedent temporal stage. Non-verbal versions of the task that rely on eye-tracking seem to suggest that children have some kind of access to their antecedent beliefs and mental representations. I just introduced a rather specific example about imagining alternatives that is highly intertwined with shrinking the minimal model. However, this shall not imply that the relevance of imagining alternatives is limited to this field. Rather, I am convinced that the ability to both anticipate probable developments of a situation and of altering the actual state of things is crucial and vastly present in human reasoning, and even more in early childhood.

In summary, I hypothesise that the introduction of alternatives of unrealised possibilities encourages to entertain parallel representations of what is likely to happen. These alternatives are more easily accessible after one event has concretely taken place, and the agents compares it with counterfactual alternatives, in order to discover how the effect could have been avoided.

3.5 Pretend play

After having introduced an account of imagination and of subtractive reasoning, I can finally attempt at tackling pretend play. As I have mentioned in chapter 2, it is virtually impossible to get any insight into the nature of pretend play without distinguishing different forms of pretence. Let me repeat the fundamental distinctions that will be used in what follows.

I assume that pretending involves an act of the imagination, which deviates from the literal representation of the world. The imagining I am after is not necessarily conscious, and it can be understood in terms of an offline simulation, akin to the one used to plan actions.

Pretending can take the form of production (of pretended actions) or understanding. The clearest phenomenal effects of pretence – e.g. carrying out certain actions that do not have literal consequences but that have imagined consequences – is characteristic of pretence production. With this term I will refer to the ability to produce certain actions that fit well with the imagined representation the child is entertaining. Even though pretence production is one of the most evident signs of pretending in children, experiments have tested separately children's ability to understand a non-literal action that is carried out by an experimenter or by a relative of the child. In real life situations, it is hard to identify situations where the child is supposed to only understand a pretend transformation or action carried out by someone else. In the case of social pretence, it is generally the case that an adult or elder mate carries out a make-believe action and prompts the child to display her understanding of the action. After the participants to the pretence have checked that they share a common make-believe settings, the child produces some action within the pretend play.

It is therefore useful to master the theoretical distinction between understanding and producing pretend play, if nothing else because it helps us identifying different abilities that a child must master in order to successfully engage in pretending. Furthermore, it seems that pretence production (at least in social contexts) requires at least a share of pretence understanding, for it seems that a child wouldn't be able to pretend in accordance with someone else pretence that φ if she didn't understand how φ may affect the world.

A pretend play has *social* nature if it involves and it's triggered by the interaction with other human beings. For instance, two siblings pretending that the bench on which they are sitting is a train is an example of social pretend play. Similarly, episodes of pretence where a parent of the child prompts the pretence is social[11]. On the opposite side, an episode of pretend play is called *individual* if there is only one agent who brings about pretend actions and behaviours, even though someone else might be watching and understanding the make-believe play.

Finally, by pretend play I will mean a play in which the child uses an object to stand for another, non-present, imaginary target or a modified version of it (see Baron-Cohen [4]). Among some authors, make-believe plays have been called properly symbolic or pretended only if the child uses a prop in a non-literal way. This latter definition of pretending excludes functional play. An instance of functional play involves a tool or prop used appropriately to its general and literal function. I am convinced that the latter definition is too restrictive: I want to be able to say that a child feeding her doll with an (empty) spoon from an empty dish is an instance of pretend play. Therefore, I will use the first definition of pretend, which includes both symbolic and functional play among instances of pretending. The distinction between functional, sensi-motor and symbolic play however will be maintained and will be used when necessary.

Now that these distinctions have been refreshed, I can introduce the various elements necessary for a pretend play. In doing this, I will sometimes draw parallelisms and contrasts with Nichols and Stich's, and with Harris' account of pretence(see Nichols and Stich [85] and Harris [43]).

3.5.1 Pretence stipulation

An episode of pretence starts with a stipulation, which establishes that some fact non-literally true now holds. The pretence stipulation can be explicit or implicit, verbally or non-verbally expressed. The agents involved in the pretend play are able to change the pretence stipulation at any later point within the play. For example, when pretence understanding is tested through non-verbal tasks, the pretence premise is non-verbally displayed to the children. Instead, two siblings playing together and pretending that the bench is a train may need to verbally and explicitly express the premise. A mother that "pours" from an empty teapot into a cup, picks up the cup and hands it to her child saying "Honey, can you serve some tea to Dolly (a doll)?", is both verbally and non-verbally starting off the pretend play, but she is not explicitly and verbally

[11] Recent studies underline the importance of the social context on pretending. Hence, it appears that social pretending shapes children's pretence. Jarrold and Conn [57].

stating the pretence premise. Rather, the child needs to figure out what the mother is stipulating in order to make sense of what she is asked to do.

An essential dimension of pretend play is that a pretence premise – i.e. whatever starts the pretence – asks for a non-literal processing of the world. Harris' theory of pretend play captures this intuition:

> [...] children do two things: having adopted a pretend stance, they suspend processing the actions and utterances of their partner according to normal rules. In particular, they stop scanning the immediate environment for situations that literally fit the utterances being produced or ways to comply literally with the requests that are being made. They recognise that a special mode of processing is called for in which the situation implied or requested by their partner is to be constructed through pretence. In addition, children compose mental "flags" or reminders, encoding the information implied in the partner's various stipulations(see [43, p. 22]).

I propose to think of the non-literal processing of the environment as the engaging in the imagination of an alternative situation. This, quite clearly, entails that not only agents are sometimes able to intervene on their imaginative activity, but also external sources may prompt and invite to engage in a certain alternative representation of the world. The "flags" mentioned by Paul Harris can be understood in a twofold manner: firstly, the pretence stipulation ϕ opens a new mental "file", which is simply based on imagining that ϕ; secondly, events and facts in the literal interpretation of the world are imported within the pretence representation on the basis of this flags. In the latter sense, pretence stipulations really much work as flags: once we have stipulated that ϕ, the information literally processed by the agent can be incorporated in the pretence if it can be reconciled with the story within the pretence.

The child is not only able to adopt a non-literal stance, but is also able to interpret the stipulations made or implied by their pretend partners. In the episode where the mum serves (non-existent) tea from the teapot, the child must be able to interpret her actions as suggesting that the mum is stipulating that there is some tea in the teapot. If this succeeds, then also further actions or requests coming from the mother will be interpreted on the basis of the stipulation "there is some tea in tea pot" which entails that there is tea in the cup, and that the doll Dolly can drink some tea from the cup. As it can be seen from this example, the processing of the pretend circumstances relies on the child's general knowledge of the world and about her ability to reason about events. This means that the child is able to apply her causal and general knowledge to the make-believe events. This idea may be expressed with the following rule:

> To understand the consequences of a pretend action, assume that the entities or substances whose existence is stipulated in the flagged information are subject to the same causal principles as their real-world equivalents (see [43, p. 23]).

I will develop some intuitions about the background theory and the causal knowledge in the subsequent sections.

3.5.2 Setting the circumstances

In my analysis of imagination of alternatives, much attention has been focused on the construction of the appropriate circumstances for the imagined event or fact. Since pretend play is a type of imagination, it still holds true that being able to pretend that ϕ does not necessarily fix a unique minimal context in which the pretend stipulation is placed. On the contrary, I can pretend that I am sitting on a train and that hence I see some poetic landscape from my window, or that I am sitting in a train and I am still (with the train) in the MoL Room. The second type of pretence is intentionally contrary to what would be generally the case if one is sitting on a train, but it is legitimate. Humans creativity and freedom of directing parts of our mental life can bring us to decide to pretend that ϕ in a highly atypical settings (or even in an impossible context). However, for the sake of formal simplicity, I will treat the setting of an impossible scenario for a pretence stipulation as a further pretence premise (or alteration of reality).

In the case of individual pretence, the child constructs the circumstances for her pretence premise in complete freedom: she may both construct a typical or most common scenario on the basis of the scripts she owns on that type of situation, or decide to deviate from the most common scenarios. She also may have to chose between different scenarios that are similarly possible.

In the case of social pretence, the participants collectively construct the scenario in which to locate the pretence premise. Therefore, the process of setting the circumstances is usually more evident and explicit, because participants need to agree on what they are about to play. For this reason, I consider episodes of social pretence as highly instructive on the processes guiding our setting the circumstances for a pretend play.

3.5.3 Background theory

In order to understand the consequences of a pretence transformation or action, the child needs to apply her general knowledge – which may take the form of a typical script of a certain situation – about the world to the events happening within the play. For instance, if she knows that phones are used to make calls, then when pretending that a banana is a phone she will pretend to make a call to someone. In what follows, I will try to explain how she can apply her general knowledge to events happening within the pretence. The explanation of how a pretended action is originated will need to await for chapter 3.5.4. Through the basic ingredients previously introduced – imagination of alternatives, general knowledge about the world, comparison between the representation of reality and the representation of the play – I will attempt to explain both intuitively –within this chapter–, and formally – in chapter 4– pretend play as a composite notion.

I have previously argued that children before age of 12 are imperfect counterfactual reasoners, and following Mackie's theory subtractive reasoning grounds our causal knowledge. However, it seems that already at an early stage children own some type of general causal general rule. I suggest that the understanding of causality starts at a early stage in the development, but it is mastered only at a mature level. Therefore, when children are able to do pretend play, they have *some* causal knowledge, which is not yet perfect. Only when we become proficient subtractive reasoners, it can be safely assumed that the ability to identify causal links is sufficiently complete.

In the chapter A.3, I have introduced the basic ideas behind the Event Calculus. Roughly, it captures the intuitive principle of inertia :

> A property persists unless it is forced to change by an event ([115, p.42])

Moreover, in the formal counterpart of this section (chap. 4.5), I will use some rules and scripts that children appear to be familiar with such as "If tea is given to someone and nothing is abnormal, tea is drunk by someone", "If X is a phone and nothing funny is going on, then X is used to call someone", "If X is food and I am hungry and nothing funny is going on, then I eat X", "when X goes to eat out, and nothing funny is going on, X arrives to the place where she will eat, orders some food, gets the food, eats the food, and leaves"[12] *et cetera*. This does not mean that I assume that all children of age between 2 and 5 share the same set of rules or knowledge about the world, and that the rules that I will be generally using can be safely assumed for all kids of that age. The knowledge of the child in question is deeply intertwined with her personal history and development, so I will be using these rules as an idealisation of what is observed in experiments and by parents.

In agreement with Harris' suggestion that a child is able to unfold a chain of events in a make-believe play on the basis of her application of her knowledge to the entities (and stipulations) in the pretence, I will assume that children are able to apply their general knowledge to pretend play, even when they are not able to relate reality and fantasy. I feel justified in drawing on this assumption, because it seems clear to me that children wouldn't be able to engage in pretend play without having some guidelines of how to go about. Furthermore, the assumption that kids are able to access their general knowledge base does not restrict on their ability to intentionally alter the rules that guide reality as they know it. Quite on the contrary, I will try to propose a formalism that accounts for both pretend plays contradicting real facts and ones that alter general rules. At this point, it should be clear to where to find the formal notions sketched in this section (chapter 4.5).

3.5.4 Pretence behaviour

Up to this point, the "file" representing the imagined alternative(s) and the one signifying reality have been kept distinct. The *rationale* for this separations resides in the consideration that following or imagining some make-believe situation, with events or transformations happening within it, does not require a comparison with reality. For me to be able to imagine that the

[12] Formally, scripts will be formalised by scenarios. For the formal definition, check appendix A.3.1.

MoL Room is a spaceship, I do not need to consciously think that the MoL Room is not *really* a spaceship or to decide whether I am now in the MoL Room, in a spaceship or anywhere else. The possibility of engaging in such considerations comes *after* my ability to imagine that the MoL Room is a spaceship. Following this train of thought, I allow imagination to include a situation that is literally the same at the currently experienced. It seems to me that only because I am typing (and I know that I am typing) it doesn't become for me impossible to imagine myself typing right-here, right-now. In chapter 4.2, the formal definition of alteration will display how I conceive the possibility of imagining situations that are literally the same as the one experienced, i.e. of a trivial alteration.

In the case of pretending, differently from the general ability to imagine a situation alternative to the one currently experienced, i.e. counterfactual imagination, the relationship between reality and pretence is somehow more involved. I will argue that, even though it is unnecessarily strong to require a conscious conceptualisation of something as "real" or "pretended", the comparison between reality and pretence plays an essential role in the appropriate production of pretend behaviour (and understanding).

A child is able to imagine what the experimenter is pretending, to follow the causal chain of events represented in from of her is understanding what happens in the pretence, only if she understands that the actions carried out do not have the same effects as ordinary actions. In order to appreciate this point it may be handy to consider an example. Suppose that an experimenter or a caretaker engages in a pretend play with a child, eliciting her to treat a Lego brick as the soap to wash up a teddy-bear, and later on that night the child tries to use a Lego brick to wash up herself. In most cases, we would say that the child hasn't been pretending that the brick is soap, but rather that she misunderstood the caretaker's or experimenter's actions. Notice, *en passant*, that this example already underlines two important dimensions for make-believe play: it is temporally bounded, and the consequences of actions within it are different from real actions. For instance, pretending to type has different consequences from typing.

The role that I have been attributing to the comparison between reality and pretence is a restricted version of the reality/pretence distinction, which only focuses on the understanding that the pretend actions and events have certain effects versus their counterpart real actions have real counterpart consequences. I do not expect children to have a conscious and linguistic awareness of what is real and how pretence differs from reality. Some studies on young children' usage of the words "real", "really" and "pretended" suggest that the prominent manner in which children employ these expressions is to express *authenticity*, rather than ontological status. This entails that if a pretended transformation looks quite authentic, children would be prone to say that it is real. However, a child pretending with her mother that the banana is a phone and who expects her mother to always call people with the banana-phone is not understanding the pretend play. One could say that the child is not playing at all!

In summary, there are couple of necessary characteristics that pretend play must satisfy in order to be considered such:

1. it must have boundaries in time, i.e. a pretence episode lasts for a limited amount of time;

2. pretended actions have different consequences than real actions (and the child has to be able to grasp this)[13];

The first characteristics suggests that pretended events and actions should be formally represented as an interval of time, if at all. This formally means that the actions and events taking place within a pretend episode will be treated as fluents – i.e. intuitively functions of time, that are treated as objects of our language – and will be conceptualised as time-dependent properties. The second feature undermines that idea that pretence understanding does not depend on kids' ability to discern reality and pretence. Since pretence comprehension has been mostly tested without controlling whether children are able to discern reality from pretence, in appendix B I sketch an extension of experiments carried out in Kavanaugh and Harris [62] to test pretence comprehension. The task I propose extends the non-verbal settings employed by Harris and Kavanaugh with a condition that discerns between reality and appearance. The intuitive idea behind this extension is that if a child understands that the experimenter is pretending that the banana is a phone and to have a phone call, she would not expect someone else – external to the spatio-temporal boundaries of the pretend play – to pick up a banana phone and have a lovely conversation with the experimenter. Even though the laws of causality and of reality may remain the same in a pretend episode, what happens in a pretence doesn't have the same consequences as it would have in reality. Symmetrically, if I now proposed you to pretend that there is a burglar in the basement, and I dialled 112 to call the police, you would probably think that I pretend quite weirdly or that I am completely bonkers.

On the basis of what I have been saying so far, pretence behaviour – i.e. the appropriate production and understanding of pretend play, the final result of all the elements previously mentioned – requires a connection and comparison of pretence with reality. In the case of pretence production, this is at best displayed by the fact that we do not literally behave as if our pretence premise were true. As Nichols and Stich [85] rightly underline, we behave in a similar way to the one would behave if our pretence premise were true. On this basis, I disagree with Harris' distinction between "counterfactual" reasoning and pretence: I actually agree about the

[13]The distinction between real consequences and pretended consequences is not meant to be all-or-nothing. For instance, I do not intend to claim that actions can only be labelled as fully "real" or fully "pretended", without intermediate levels. Rather, I suggest that the production of pretence is only possible on the basis of some comparison between the literal interpretation of the world and an imaginative stance on utterances and actions. On the basis of the experimental results in the counterfactual tasks and in the False Belief Tasks, I hypothesise that children start grasping the distinction between intended consequences and literal consequences of a pretended action between 3 and 4 years. As previously mentioned, the theory here developed is meant to characterise the reasoning processes displayed by children around this age, therefore it is acceptable to assume that children will display some understanding of the difference between real and pretended actions around this age (in agreement in the experimental results in Kavanaugh and Harris [62]). However, I do not hypothesise that children 3 or 4 years old perfectly discern reality and pretence as adults do.

distinction, but my story slightly diverges on how it should be drawn. Paul Harris [43] argues that "counterfactual" reasoning differs from pretend play because the former is set up as a departure form reality, whereas the latter are essentially self-contained and independent from reality.

> [T]he type of suppositional thinking that is deployed in pretend play is importantly different from the counterfactual thinking that may be directed at reality itself. In the case of pretend play, children do not set up a contrast between an imaginary event and an actual event. They simply invent, watch, or describe an imaginary event that has no close cousin in reality with which it contrasts [. . .] Pretend events are essentially self-contained; they are not set up as departures from actual events. Indeed, consistent with this analysis, a game of pretend can turn on a situation that happens to coincide with the way things really are [. . .] ([43, p. 124])

Both Harris and I claim that a game of pretence can be about a situation that is exactly the same as the real one. However, I affirm that in order to output a proper pretend behaviour a child needs to distinguish between real actions (and their consequences) and a pretend action. Within my story, this is only possible if the child is able to reason at the same time about pretence and reality. This means that the "reality file" or program ends up representing both how the environment literally appears to the agent and that she is pretending something. The child representation that she is pretending ϕ requires an import from the simulated or imagined situation to the normal representation of reality, that guides actions and behaviours. I do not mean to claim that children have an explicit and conscious representation that they are pretending, this can be totally unbeknown to them. But because I want to explain pretending in terms of our simpler abilities, I cannot see a way in which acting according to a pretend play can be independent from reality. At the end, imagining that there is a monster under my bed doesn't make me run out of my room, call my friends, and so on and so forth even though it triggers certain emotional reactions. In the case of pretence understanding, in order to report to the experiment which one is outcome of the pretended transformation, the child must be able to act upon her beliefs and imagined representations.

3.6 Autism: some hypothesis

As I disclosed in chapter 2.6.3, the main theoretical hypothesis grounding the present work is that the *dissociation hypothesis* generalises to pretend play. Robust support has been found (see Smid [107],van Lambalgen and Smid [116], van Lambalgen and Stenning [117] and Pijnacker et al. [92]) for the idea that subjects with autism are unimpaired relative to controls in Closed World Reasoning about Rules (CWRr), at least in logical inference tasks. At the same time, the results obtained clearly support the hypothesis that people with autism have problems handling exceptions to rules, i.e. closed world reasoning about facts.

The dissociation hypothesis was originally formulated on the basis of logical inferences tasks (e.g. Modus Ponens, Modus Tollens, Affirmation of the Consequent *et cetera*). In what follows (and in chapter 4), I will attempt to display how this distinction generalises to the case of pretend play. In order to appreciate this, it becomes essential to make explicit that Closed World Reasoning about rules will be formally represented by integrity constraints (IC). IC were initially proposed by Kowalski [66, 33] to model abductive reasoning in Logic Programming. An introduction to abduction in Logic Programmin can be found in chapter A.4. Informally, integrity constraints are always based on a given program P, and hence capture the intuition that abductive explanation depends on the rules I (even implicitly) entertain about the *explanandum*. Closed World Reasoning about facts, instead, is formally represented by predicates or formulas representing abnormalities and exceptions to a rule. In the propositional case, a rule in the logic for planning used will have the following form:

$$\psi \wedge \neg ab \rightarrow \varphi$$

where ab stands for an abnormality, which may be triggered by certain situations $e_1, \ldots, e_n$. What triggers the abnormality is defined in a program P with the clauses

$$e_1 \rightarrow ab,$$
$$\ldots,$$
$$e_n \rightarrow ab$$

The reader might get the impression that the hypothesis proposed is really much intertwined with the logic: this is partially motivated by the fact that I conceive chapters 3 and 4 as complementary in providing a theory of pretence. In any case, the formulas just introduced are only meant to visually display how one can understand exceptions. I will henceforth attempt at providing an explanation of how CWRf and CWRr affect children with autism pretence production and understanding as informally as possible.

CWRf plays a fundamental role in the initial steps of a pretend play episode, for a child needs to inhibit the *prepotent*, literal interpretation of the world in order to be able to switch to the "pretence file". More precisely, children will need to relate (at least) two incompatible facts about the world, and to recognise that they cannot be simultaneously true. For instance, in the case of the empty tea-pot, the child will represent that the tea-pot *is* empty even when it is grabbed, moved as if one was pouring liquid from it into a mug, and the mug is offered to drink some tea. It will hence be necessary that the child handles this contrasting information. On the basis of clues (e.g. no liquid actually comes out of the tea-pot), the child will understand that the tea-pot is used as-if it was full. Therefore, she will move to a pretend environment, where "the tea-pot is full of tea" is added as a tag. However, assuming that subjects with autism suffer from an impairment in exception-handling (or CWRf), difficulties may arise already at the level of the pretence stipulation (and hence pretence understanding), in some of the cases where the latter requires to deal with abnormalities[14].

The construction of a minimal model making true what's happening in the pretence – i.e. representing the "pretence file" –, as it was hinted in section 3.3.1, are modelled in terms of abduction, i.e. Closed World Reasoning about Rules. Once the child has started the "pretence file", the generation of what must be true in order for the pretence premise to hold true is obtained on the basis of the rules that define the pretence premise and that the child has in her "mental database". Since according to the dissociation hypothesis, children with autism are unimpaired in CWRr, this means that the theory suggests that once the child has accepted the pretence premise the generation of its minimal model will not be problematic. Notice that this fits well with the results claiming that prompted and scaffolded pretence production are relatively unimpaired for autistic subjects.

I have claimed that the generation of a minimal model representing what's happening in the pretend play – according to the dissociation hypothesis – won't be problematic for children with autism. This claim should be clarified as follows: even though the generation of the pretence file should not cause difficulties, it might happen that the CWRf impairment will affect the construction of the circumstances. If children with autism have problems handling abnormalities, they may be unable in many cases to inhibit a prepotent response or to "generate" novel exceptions that would keep the pretend play going. To give but an example, consider the following case: a child with autism pretends that the tea-pot is full. She pretends to serve some tea and then she pretends to drink it. The pretend episode terminates. She might have pretended to offer the tea to a doll, or pretended to spill some tea in order to then clean it up, she may drink the tea and then pretend to prepare more tea . . . However, any of these actions requires the realisation that the simple script "Someone drinks tea if she is given tea, she brings it to the mouth and takes sips of it." can be interrupted and altered by many exceptions. As unclear as it may sound, I ask the reader to trust the intuitions just sketched until the end of chapter 4 (or to directly read section 4.6), where I formally display how these ideas originate from the hypothesis itself and from the logic used.

Finally, both elicitation and scaffolding "force" the subject to modify their representation of the world to satisfy the prompt or scaffolding. In terms of a logic program this corresponds to saying that the program representing reality must be changed to satisfy an integrity constraint. Therefore, I will make use of the latter to formalise the former. The analogy between integrity constraint and elicitation is not merely informal, rather, it is a request to carry out a certain

[14]The reader may wonder why I didn't say in "all" cases where what is true in realty and what is entailed by the pretend actions are two incompatible things. It is very well possible that in all such cases children with autism will display some difficulty in accepting the pretended actions. However, as it will be further displayed in chapter 4.6, the dissociation hypothesis is not "all-or-nothing". The dissociation hypothesis doesn't claim that people with autism are never able to handle abnormalities (it will be really hard to explain how they carry out on structured and targeted action in general), but rather that they display an *impairment*. For a more detail explanation of this point, the reader can have a look to footnote 14 on page 87. For the present aims, it suffices to say that in most (but not all) cases people with autism seem to struggle in handling exceptions.

pretend action or to assume a certain pretence premise is precisely symmetric to finding out about some fact φ and engaging in abductive reasoning about it. In both cases, something is forced upon the minimal model of the subject, and the forced information must be made consistent and to be explained. As the reader might guess, IC are a form of CWRr and the latter is unimpaired in autism according to the working hypothesis here used. Therefore, consistently with the empirical results, the theory I am advancing suggests that elicited pretend place will facilitate children with autism' pretend play, by forcing a resolution to the conflict between contrasting facts that the child was unable to resolve herself.

3.7 Conclusion

In this chapter, I have proposed an highly theoretical story on how counterfactual imagination, subtractive and counterfactual reasoning, and pretend play are related to each other. Counterfactual imagination has been treated as the spring from which subjunctive reasoning and pretend play originate. I have furthermore displayed why subtractive reasoning predates full-fledged counterfactual reasoning, understood as the "sense for the nearest possible world". The latter appears to be more cognitively requiring than the imagination of what would have been if an event e didn't take place, imagination guided by simple conditional rules.

Sections 3.5 and 3.6 have displayed how the theoretical story I proposed applies to the case of pretend play, both for neuro-typical and autistic children. Within section 3.6, the dissociation hypothesis has been roughly correlated with the different phases of a pretend play. A first approximation of the analysis of pretence in terms of CWRf and CWRr appeared to correspond nicely to many empirical results about autistic performance in pretence understanding and production tasks. Furthermore, this approximation generated a novel hypothesis on the causes of the so-called "generativity" deficit in autism: *contra* Jarrols et al. [59], I suggested that the repetition of stereotypical scripts and the shorter engagement in pretend production can be explained on the basis of the dissociation hypothesis.

As previously mentioned, both the hypothesis which grounds the thesis and the story I want to advance are highly dependent on the logical formalisation of the theoretical ideas sketched in this chapter. The following chapter will be devoted to the formalisations of counterfactual imagination, subtractive reasoning and pretend play.

CHAPTER 4

A LOGICAL FORMALISATION OF PRETENCE

Following in the footsteps' the structure of chapter 3, this chapter will introduce the formal counterparts of the notions that constituted the building blocks of my story on imagination, pretence and subtractive reasoning.

For the sake of readability, I am going to introduce some formal notions without dwelling upon all the formal details. The reader without any familiarity with logic programming, negation as failure and the event calculus will find some introduction to these topics in appendix A, with references to further literature. The main logical results about the formalism used as the basis for this chapter have been proved in van Lambalgen and Hamm [115], Stuckley [110], and they are introduced in app. A.

4.1 Introduction

I am going to use a first order language to mimic the way in which we represent our surroundings. The language here used however does not simply represent the fact that an object standing in front of you is a banana by $B(a)$ where a is the object standing in front of you, and the predicate B denotes the property of being a banana. Rather, a many sorted logic will be needed, which allows predicates to be indexed by time and space. The usage of a multi-sorted logic, in conjunction with the Event Calculus, will allow me to express that properties apply to objects at a certain time and they can be affected by events – e.g. moving a banana will affect the spatio-temporal location at which the agent sees certain features and hence at which she identifies a certain type of object.

The following list enumerates the five sorts of objects that the logic will consider:

1. a set of individual objects O;

2. a set of tuples of real numbers representing space S;

3. real numbers representing time T;
4. time-dependent properties P;
5. variable quantities Q;
6. event types, whose instantiations (tokens) typically mark the beginning and end of time-dependent properties E.

Notice that while properties are, in general, represented by predicates, in this case the properties that vary over time are objects of the language, able to be the argument of a predicate. I will use the terms "reification" and "nominalisation" to refer to the treatment of properties as objects, connecting two traditions in artificial intelligence and linguistics. Among the motivations to adopt this peculiar stance on the status of properties, consider an example of causation and accomplishment. The accomplishment "switching on the light" has an internal structure: in order to say that I managed to switch on the light, it is necessary to express that an object having a certain time-dependent property such as "at time t the light is off" and "at time t the switch is in the downward position", can lose or acquire properties on the basis of certain actions that I carry out such as going towards the switch, moving the switch from downward to upward position at time t'. Finally, I want to be able to express that the property of the switch being in downwards position at time t is modified by my action of moving it to the upward position at time t'. Hence, it will be required to be able to express time-dependent properties as objects in our language, that are able of filling slots in the predicates.

The choice of using spatial intervals and time-dependent properties is intended to mimic the manner in which objects are recognised and treated on the basis of their features, which are co-located in space and time. Following the work of A. Treisman (Treisman [112],Treisman and Gelade [113],Treisman and Schmidt [114]), following the intuition that "[f]eatures which are registered in the same location and temporal interval, i.e., within the same central focus of attention, can then be coded as belonging to the same object." (see [112, p. 1]). Furthermore, as displayed in Landau et al. [68],Deák et al. [22], it seems that attentive visual features of object play an important role in determining the function of the object. This means that the event calculus – as presented in van Lambalgen and Hamm [115] – and the modification of it introduced in appendix A represent how objects are independent from spatial and temporal locations, while their prototypical features are not[1]. Henceforth, I will use the word "object" to refer to individual objects, and I will treat the prototypical features characterising an object at a spatio-temporal location as fluents.

Spatio-temporally dependent properties will be called *fluents*: a property that depends on space and time locations can be informally understood as what is initiated and ceased by events.

[1] Formally, features of objects are going to be treated as fluents, i.e. functions from time to truth values. This means that the characteristic features of an object, e.g. shape, will be defined by a fluent which has an additional argument for the object, and the spatial location.

Hence, fluents will be represented as objects of the language, even though it is suggested to think of them as functions of time. However, when a fluent d contains an additional parameter x, it will be useful to think of $d(x)$ as a function that maps x to a fluent-object. In section A.3.6 it will be explained how fluents with additional parameters can be treated as proper objects.

Within this chapter, I will assume some acquaintance with the Event Calculus and with Logic Programming. The details of the system I have in mind – which is based on the work of van Lambalgen and Hamm [115] – are laid out in appendix A. The Event Calculus can be thought of as a logical system that tries to capture the intuitive principle of inertia, according to which change is always due to a cause. The system called Event Calculus (cfr. Kowalski and Sergot [66], Shanahan [105] and Miller and Shanahan [82]) captures two notions of causality or change:

1. instantaneous change, e.g. collision,

2. continuous change, e.g. acceleration due to the gravitational field.

For clarity's sake, among the representations that an agent entertains during an interval of time – i.e. a logic program $\mathscr{P}$ – I am going to distinguish between a Knowledge Base (KB) and some facts (P). The Knowledge Base is meant to include the most general rules about the world. For instance, rules about causality and rules concerning general types of situations (or scenarios) will be included in KB. The set of facts P will represent the context-specific representations the agent entertains, so for instance facts about the world and exceptions to a rule may be specified only in P.

The plan for the chapter, unsurprisingly, is symmetric to the structure of chapter 3. I am going to start off by characterising, in section 4.2, how a imagined alternative can be obtained from the representation of reality. This will be done in terms of a function, which substitutes formulas with other formulas. On the basis of this function, I am going to obtain an "alternative", i.e. the formula which defines the representation of what is going on in imagining or pretending. In sections 4.2 and 4.3, I am going to display the difference, in logical form, of *imagining something* (φ) from the construction of the circumstances in which the φ is imagined.

Within section 4.4, I am going to display how subtractive reasoning differs from full-fledged counterfactual reasoning. I am going to model some examples of counterfactual tasks taken from the literature. These model are going to prove that, according to the proposed story, full-fledged counterfactual reasoning – i.e. the "sense for the nearest possible world" described in Perner et al. [88] – is more complex than subtractive reasoning based on simple conditional rules.

The chapter will find its close with the formalisation of pretend play. In section 4.5, I am going to display how an alteration from the reality program and the construction of its circumstances apply to the case of pretence. Furthermore, in section 4.5.2 I am going to advance a model for the most complex type of pretend play, i.e., object substitution. Section 4.6 will conclude the chapter, by displaying how the formalisation of pretend play advanced applies to the case of children with autism. With the aim of relating the present approach with the experimental results on pretending in autism, I am going to indulge on elicited pretence.

4.2 Imagining alternatives, formally

I will now proceed to formally define what it means to imagine counterfactually, i.e. to imagine something that is different from reality. On the basis of the distinction drawn in chapter 3 between imagining an alternative φ and constructing some minimal circumstances s.t. φ is true, I will start off defining what it means to imagine an alternative with regards to our language.

Definition 4.1. Let ψ, ψ' be two formulas in the set of all formulas $Form$ be fixed. A function $\sigma_{\psi,\psi'}$ is a *replacement function* if $\sigma : Form \rightarrow Form,\ \chi \mapsto \chi_{[\psi/\psi']}$ for any $\chi \in Form$. We omit the subscript of σ if it is clear from the context.

Definition 4.2. Let ψ, ψ' be fixed. We say that a replacement function $\sigma_{\psi,\psi'}$ *respects* $\mathscr{P}$ if $\psi \in \mathscr{P}$, and the substitution of ψ with ψ' generates a formula $\xi = \sigma_{\psi,\psi'}(\chi)$ s.t. either $\xi = \psi$ or $\xi \notin \mathscr{P}$.

Definition 4.3. (Alteration) A formula φ^A is called a $\mathscr{P}$-alteration iff $\sigma_{\psi,\psi'}$ respects $\mathscr{P}$ and $\varphi = \sigma_{\psi,\psi'}(\chi)$ for a $\chi \in Form$.

Notice that in the definition 4.3 we do not distinguish whether $\varphi \in P$ or $\varphi \in KB$. Henceforth, I will specify where the modified formula comes from, if this is not clear from the context. It is furthermore needed to keep track of the formula that are modified.

Definition 4.4. (imagination program) Let $\mathscr{P} = KB \cup P$ be a logic program. Let σ be defined over two fixed ψ, ψ'. We say that $\mathscr{P}^A = KB^A \cup P^A$ is an *imagination program with alternative set of clauses* iff $P^A := \{\varphi^A\}$ and $KB^A = KB$.
We say that $\mathscr{P}^A$ is an *imagination program with alternative knowledge base* iff $KB^A = \{\varphi^A\}$ and $P^A = P$.

As it can be read off from the definition, I do impose the restriction according to which alterations cannot affect at the same time the knowledge base and the set of clauses[2].

I will now prove that definition 4.2 suffices to model the relevant cases of imagining alternatives introduced in 3.2: object substitution, attribution of absent properties, imagining absent objects or entities.

Definition 4.5. (Object substitution) Let $\varphi = F(a)$ be in $\mathscr{P}$. We say that φ^A is an *object substitution* if $F(b)$ is in $\mathscr{P}^A$, and $F(b) \notin \mathscr{P}$.

Proposition 4.1. *The definition of $\mathscr{P}$-alteration (def. 4.3) covers the case of object substitution (def. 4.5).*

[2]In general, alternations may affect both the knowledge base and the set of clauses, but this requires a serial procedure.

Proof. Let the ψ, ψ' on which σ is defined be $\psi = F(a)$ and $\psi' = F(b)$. Then by definition 4.1, $\sigma(F(a)) = F(a)_{[F(b)/F(a)]} = F(b)$. Furthermore, by definition 4.2, either $F(a)_{F(b)/F(a)} = F(a)$ or $F(b) \notin \mathscr{P}$. Hence, $F(b) \notin \mathscr{P}$. Finally,by def. 4.3, $\varphi^A = F(b)$. ■

Definition 4.6. (Attribution of absent properties) Let φ be in $\mathscr{P}$. We say that a formula $(\varphi \wedge \chi)$ is an *attribution of absent properties* if $(\varphi \wedge \chi)$ is in $\mathscr{P}^A$, and $\chi \notin \mathscr{P}$.

Proposition 4.2. *The definition of $\mathscr{P}$ alteration (def. 4.3) covers the case of attribution of absent properties (def. 4.6).*

Proof. Let ψ, ψ' over which σ is defined be $\psi = \varphi, \psi' = \varphi \wedge \chi$. Then by def. 4.1, $\sigma(\varphi) = \varphi_{[\varphi/\varphi \wedge \chi]} = \varphi \wedge \chi$. Furthermore by def. 4.2, either $\varphi = (\varphi \wedge \chi)$ or $(\varphi \wedge \chi) \notin \mathscr{P}$. Since $\varphi \in \mathscr{P}$ by assumption, $\chi \notin \mathscr{P}$. Finally, by def. 4.3, $\sigma(\varphi) = \varphi^A$. ■

Definition 4.7. (Imagination of absent objects) Let φ be in $\mathscr{P}$. We say that a formula $\varphi \wedge \psi)$ is *an imagination of absent objects* if $(\varphi \wedge \psi)$ is in $\mathscr{P}^A$, $\psi = F(b)$ for b an object-fluent and $F(b) \notin \mathscr{P}$.

Proposition 4.3. *The definition of alteration (def. 4.3) covers the case of imagination of absent objects (def. 4.7).*

Proof. Let the ψ, ψ' over which σ is defined be $\psi = \varphi, \psi' = \varphi \wedge F(b)$. By def. 4.1, $\sigma(\varphi) = \varphi_{[\varphi \wedge F(b)/\varphi]} = \varphi \wedge F(b)$. Furthermore, by def. 4.2, either $\varphi = (\varphi \wedge F(a))$ or $(\varphi \wedge F(b)) \notin \mathscr{P}$. Since $\varphi \in \mathscr{P}$ by assumption, $F(b) \notin \mathscr{P}$. Finally, by def. 4.3, $\varphi^A = \sigma(\varphi)$. ■

The attentive reader will have noticed that the definition of object substitution and of attribution of absent properties have been formulated in terms of fluent-objects. Let me briefly indulge in a motivation for this choice, aside the formal one that the only predicates formally allowed in the language are the programmed ones of the event calculus – which intuitively should not be affected. Consider the case of attribution of a certain property, e.g. being a banana, to an object that stands in front of you. This could be represented in various manners:

1. a time-invariant formula $B(a)$, where a denotes the object in front of you and B represents the property of being a banana;

2. a time-dependent (or t-property) property, e.g. $Holdsat(banana[x], t)$;

3. a spatio-temporally dependent property. Let $Holdsat(banana, l, t)$ represent the spatio-temporally dependent property that at location l at time t a banana is recognised, hence it is true that the object banana is at a certain spatio-temporal location.

The second possibility is employed in van Lambalgen and Hamm [115], Intuitively, events and actions may affect the attribution of a t-property to x, e.g. the property of being upwards for a switch. For instance, if someone comes in and switches the switch, it will cease to have the property of being upwards and start being upwards. Within this chapter, I am going to make use of the third option, which extends the approach in van Lambalgen and Hamm [115] by modelling both spatial and temporal inertia. This means that an object doesn't change location unless it is forced to do so. Notice, *en passant*, that the latter approach is conceptually more appropriate to represent the manner in which objects are dependent on their prototypical features of features. Intuitively, the objects in my visual field are recognised *as* a certain instantiation of an object on the basis of a complex process of parallel processing of features, which are then compared with object-types (see Treisman [112]). Furthermore, the manifold of features of the object I recognise may vary its position over time. Suppose that the object in front of me is a laptop, i.e. running though its characteristic dimensions and features $\psi_1, \ldots, \psi_n$ I recognise it as such. If later the laptop is moved I need to compute that the new instantiation of the features $\psi_1, \ldots, \psi_n$ at a different spatio-temporal location from before is the very same object that I had recognised before. Events and actions can affect objects, i.e., the recognition of an object-type in a sensible manifold can stop because the object has been move, because some of its prototypical features cease being instantiated, *et cetera*.

In the previous chapter (i.e. chap. 3), I attempted to argue that the production of pretence appropriate behaviour is only possible on the basis of a comparison between the "reality file" and the imagined alternative. Therefore, in section 3.5.4, I will define the manner in which an imagination program (alternatively, pretend program) is imported into the "actual" representation of the world. Furthermore, I will show that the alteration defined characterises a set of formulas that is bijective to the original set of clauses.

4.3 Circumstances, formally

In chapter 3.3, I have argued for the theoretical distinction between imagining *that* something is different and imagining its circumstances. Mirroring this argument, the previous section defined an alteration of a program, and I will now formally display how the minimal circumstances for some φ^A are constructed. This will allow me to display how subtractive reasoning formally differs from counterfactual imagination and from counterfactual reasoning à la Perner et al. [88].

Now that I have defined an alteration of a program $\mathscr{P}$, it is possible to display how to provide the context to this imagined alternative.

Definition 4.8. (Minimal circumstances) Given a imagination program $\mathscr{P}^A$ and $\varphi^A \in \mathscr{P}^A$, its minimal circumstances are obtained as follows:

1. if $\mathscr{P}^A$ is an imagination program with alternative set of clauses, and
   a) If φ^A is a literal[3], and there is an unique $\chi \in P$ whose head is φ^A, then the body of χ is part of $\mathscr{P}^A$.

[3] This is, if $\varphi^A = F(a)$ for some predicate F and some constant a, or $\varphi^A = \neg\psi$ for some atom ψ

b) if $\varphi^A = (\neg)F(a,\dots) \wedge (\neg)P(a,\dots) \wedge \dots (\neg)P(b,\dots)$ and for any literal there is an unique clause $\chi \in P$ whose head is the literal, then the body of each clause is included in $\mathscr{P}^A$.

c) if $\varphi^A = \neg\psi$, and there is an unique clause $\chi \in P$ whose head is $\neg\psi$, then the body of χ is added to $\mathscr{P}^A$.

d) if $\varphi^A = \neg\psi$ and there is an unique clause $\chi \in P$ with body ξ whose head is ψ, then the negated body of χ (i.e. $\neg\xi$) is added to $\mathscr{P}^A$.

e) if $\varphi^A = \psi \vee \chi$, and for any literal α appearing in $\psi \vee \chi$ there is an unique clause in P with head α, then the body of the clause is added to $\mathscr{P}^A$.

f) if $\varphi^A = \psi \wedge \chi$, and for any literal α appearing in $\psi \wedge \chi$ there is an unique clause in P with head α, then the body of the clause is added to $\mathscr{P}^A$.

g) if φ^A is a conditional $\psi \rightarrow \chi$ and there are two unique clauses $\xi_1, \xi_2 \in P$ whose head is respectively ψ, χ, then the body of ξ_1 and the body of ξ_2 is added to $\mathscr{P}^A$.

2. if $\mathscr{P}^A$ is an imagination program with alternative knowledge base, and $\varphi^A \in KB^A$, and there are two unique clause ψ_1 and $\psi_2 \in KB$ s.t. the head of ψ_1 is the antecedent of φ^A and the head of ψ_2 is the consequent of ψ, then the bodies of ψ_1 and ψ_2 are added to $\mathscr{P}^A$.

3. Otherwise, nothing is added to $\mathscr{P}^A$.

This definition is meant to capture the intuition that if there are some minimal circumstances s.t. φ^A is true and there is no other minimal circumstance making φ^A true, then it makes sense to imagine φ^A with this minimal circumstances. However, it will mostly be the case that a formula φ^A does not have an unique clause whose head is φ^A, because in programs it is allowed to have multiple bodies entailing the same head, and this is often the case for general facts.

Following the definition, if there is no minimal condition that is able to make φ^A true (and nothing else), only φ^A is added to the program, and I consider φ^A itself as the minimal context for itself[4].

The proposed definition suffices to display how the minimal circumstances can be constructed, but it does so with a rather narrow conception of "minimal": definition 4.8 is really concerned only with the identification of those clauses that are strictly necessary and sufficient – on the basis of the agent's information and in the circumstances – for φ^A to be true (and nothing more). In general, this won't capture the manner in which some circumstances for our imagined alternatives are constructed, where we try to imagine something that fits as closely as possible the props of an experimenter or of a playmate, where we try to imagine a story about the imagined alternative and so on and so forth. Therefore, I also need to provide a general strategy for constructing these circumstances, that are still going to work with closed world reasoning about rules.

[4]Note that the definition just proposed mimics quite closely the general strategy for updating a program in the case of abduction: i.e. modifying the program P s.t. φ holds by looking at the SLD(NF) tree. However, the definition in an important sense is a simplified version of abduction, because it requires the SLDNF tree to have a simple form.

This will be done in terms of integrity constraints. Intuitively, the following definition captures the required procedure:

Definition 4.9. If $\mathscr{P}^A$ is a imagination program with alternative set of clauses, then for any $\varphi^A \in P^A$, the integrity constraint

IF $?\varphi^A$ succeeds and $\varphi^A \leftarrow {\psi_1}^A, \ldots, {\psi_n}^A$ are all the clauses (if any) with φ^A as a head in $\mathscr{P}^A$, THEN one of $?{\psi_1}^A, \ldots, \psi_n^A$ also succeeds, where ${\psi_i}^A$ is a possibly empty alteration of a $\psi_i \in \mathscr{P}$.

is added.
If $\mathscr{P}^A$ is an imagination program with alternative knowledge base, then for any $\varphi^A \in KB$ the integrity constraint

IF $\varphi^A = \psi^A \leftarrow \chi^A$ and $?\chi^A$ succeed, THEN also $?\psi^A$ succeed.

is added to the program.

More formally, what will be done is to extend σ as to generate a set of clauses that "explain" it. The following definitions are formulated for an imagination program with alternative set of clauses P^A, but they can be easily modified for an imagination program with alternative knowledge base.

Definition 4.10. (Constructing the circumstances) Let ψ, ψ' over which σ is defined be fixed. Let $\varphi \in P$, and $\varphi^A := \sigma(\varphi)$. The imagination program P^{A*} describing the circumstances for φ^A is originated from P through $\sigma^* : P \mapsto P^{A*}$, where $P^{A*} = \sigma^*[P^A]$.

On the basis of $\sigma(\phi) = \varphi^A$ as defined in 4.3, $\varphi^A \in P^A$. Then $\sigma^+ : P \mapsto P^A$, where $P^A = \sigma^+[P]$, is defined as follows for any $\chi \in P$

$$\sigma^+(\chi) = \begin{cases} nothing_\chi & \text{if } \chi_{[\varphi^A/\varphi]} = \varphi \text{ or if } \chi_{\varphi^A/\varphi} \vDash_3 \perp; \\ \chi_{[\varphi^A/\varphi]} & \text{otherwise} \end{cases}$$

Now, finally, σ^* is defined as follows for any $\chi \in P$

$$\sigma^*(\chi) = \begin{cases} \chi_{[\psi_i^A/\psi]} & \text{if } \psi_i^A \text{ is obtained from abduction over } ?\varphi^A \text{ wrt } P^A \\ & \text{and } \psi^A \text{ is the result of an alteration of a} \psi \in P,; \\ \sigma^+(\chi) & \text{otherwise, for any } \chi \in P; \end{cases}$$

Proposition 4.4. *The function σ^+ is a well defined, injective and surjective function.*

Proof. Firstly, σ^+ is a well defined function. Suppose that $\chi = \psi$, for $\psi, \chi \in P$. Then given the original substitution $[\varphi^A/\varphi]$ which constitutes the only $\varphi^A \in P^A$. To prove that σ^+ is a well defined function it suffices to show that $\sigma^+(\chi) = \sigma^+(\psi)$.

1. if $\chi_{[\varphi^A/\varphi]} = \varphi$ then also $\psi_{[\varphi^A/\varphi]} = \varphi$. Hence $\sigma^+(\chi) = nothing_\chi$ and $\sigma^+(\psi) = nothing_\psi$, and by the initial assumption that $\chi = \psi$ we obtain that $\sigma^+(\chi) = \sigma^+(\psi)$.

2. if $\chi_{[\varphi^A/\varphi]} = \bot$ then also $\psi_{[\varphi^A/\varphi]} = \bot$. Hence $\sigma^+(\chi) = nothing_\chi$ and $\sigma^+(\psi) = nothing_\psi$, and by the initial assumption that $\chi = \psi$ we obtain that $\sigma^+(\chi) = \sigma^+(\psi)$.

3. otherwise, $\sigma^+(\chi) = \chi_{[\varphi^A/\varphi]} = \psi_{[\varphi^A/\varphi]} = \sigma^+$.

Secondly, I want to show that σ^+ is injective, i.e. if $\chi \neq \psi$ then $\sigma^+(\chi) \neq \sigma^+(\psi)$.

1. if $\chi_{[\varphi^A/\varphi]} = \varphi$ then $\sigma^+(\chi) = nothing_\chi$.

   a) if $\psi_{[\varphi^A/\varphi]} = \varphi$ or $\psi_{[\varphi^A/\varphi]} = \bot$ then $\sigma^+(\psi) = nothing_\psi$ and by assumption that $\psi \neq \chi$, $nothing_\chi \neq nothing_\psi$.

   b) suppose that $\psi_{[\varphi^A/\varphi]} = \chi$. By def 4.3, the substitution over atoms cannot substitute predicates nor terms that are already in the program P. Hence, if $\psi_{[\varphi^A/\varphi]} = \chi$ then this can only be because the substitution carried over φ was $[\varphi/\varphi]$ and hence $\psi = \chi$ contradicting our initial assumption.

2. if $\chi_{[\varphi^A/\varphi]} = \bot$ then $\sigma^+(\chi) = nothing_\chi$ and

   a) if $\psi_{[\varphi^A/\varphi]} = \varphi$ or $\psi_{[\varphi^A/\varphi]} = \bot$ then $\sigma^+(\psi) = nothing_\psi$ and by assumption $\psi \neq \chi$.

   b) suppose that $\psi_{[\varphi^A/\varphi]} = \chi$. By def 4.3 (3), the substitution over atoms cannot substitute predicates nor terms that are already in the program P. Hence, if $\psi_{[\varphi^A/\varphi]} = \chi$ then this can only be because the substitution carried over φ was $[\varphi/\varphi]$ and hence $\psi = \chi$ contradicting our initial assumption.

3. otherwise $\sigma^+(\chi) = \chi_{\varphi^A/\varphi}$ and $\sigma^+ = \psi_{\varphi^A/\varphi}$. By def. 4.3 and 4.1, $\psi_{\varphi^A/\varphi} \neq \chi_{[\varphi^A/\varphi]}$.

Finally, σ^+ is obviously surjective by definition of P^A. ■

Proposition 4.5. *The function σ^* defines a bijection between P^{A*} and P.*

Proof. Firstly, I need to show that σ^* is a well defined function. To do this, it suffices to show that if $\chi = \xi$, for $\chi, \xi \in P$, then $\sigma^*(\varphi) = \sigma^*(\psi)$.

1. Suppose $\sigma^*(\chi) = \chi_{[\psi^A/\psi]}$ where ψ^A is obtained from abduction over $?\varphi^A$ and $\sigma(\psi) = \psi^A$ for a $\psi \in P$. Then because $\chi = \xi$, also $\sigma^*(\xi) = \xi_{[\psi^A/\psi]} = \chi_{[\psi^A/\psi]} = \sigma^*(\chi)$.

2. otherwise, $\sigma^*(\chi) = \sigma^+(\chi)$ which by prop. 4.4 and $\chi = \xi$ is equal to $\sigma^+(\xi) = \sigma^*(\chi)$.

Secondly, I need to show that σ^* is injective, i.e. that if $\chi \neq \xi$ for $\chi, \xi \in P$, then $\sigma^*(\chi) \neq \sigma^*(\xi)$.

1. If $\sigma^*(\chi) = \chi_{[\psi^A/\psi]}$, and

   a) $\sigma^*(\xi) = \xi_{[\psi^A/\psi]}$, then by definition 4.1 and 4.3 $\xi_{[\psi^A/\psi]} \neq \chi_{[\psi^A/\psi]}$, otherwise some formula substituted would have to be already in P.

b) $\sigma^*(\xi) = \sigma^+(\xi)$, then $\sigma^+(\xi)$ cannot be equal to $\sigma^{[\psi^A/\psi]}$ otherwise the abduction wouldn't have produced ψ^A.

2. If $\sigma^*(\chi) = \sigma^+(\chi)$, and

a) $\sigma^*(\xi) = \xi_{[\psi^A/\psi]}$ for ψ^A generated through abduction over $?\varphi^A$, then $\sigma^*(\xi)$ cannot be equal to $\sigma^+(\chi)$ otherwise the abduction wouldn't have produced ψ^A.

b) $\sigma^*(\xi) = \sigma^+(\xi)$ then by injectivity of σ^+ we obtain that $\xi_{[\psi^A/\psi]} \neq chi_{\psi^A/\psi}$ and hence $\sigma^*(\xi) \neq \sigma^*(\chi)$.

Third, I need to show that σ^* is surjective, i.e. $\sigma^*[P] = P^{A*}$ where $\sigma^*[P] = \{\sigma^*(\chi) \mid \chi \in P\}$. This follows immediately from the definition of P^{A*}.

■

The fact that σ^* is bijective, means that there is a unique inverted function $\sigma^{*-1} : P^{A*} \to A$, which allow one to derive back P from P^{A*}. This result can be strengthen showing that the two minimal models of P and P^{A*} are isomorphic.

Proposition 4.6. *The minimal model of a program $T(P)$ and of an altered program $T(P^{A*})$ s.t. $\sigma^*(P) = P^{A*}$ are isomorphic in the following sense: σ^* is a bijection between P and P^{A*} and*

1. *For any predicate latter ${A^n}_j \in P$ and for any $b_1, \ldots b_n$ in P, $T(P) \models_3 {A_j}^n(b_1, \ldots, b_n)$ iff $T(P^{A*} \models {A^n}_j[\sigma^*(b_1, \ldots \sigma^*(b_n]$;*

2. *for any function letter ${f^n}_j$ in P and for any $b_1, \ldots, b_n$ in P, $T(P) \models {f^j}_n(b_1, \ldots, b_n$ iff $T(P^{A*}) \models {f^j}_n[\sigma^*(b_1), \ldots, \sigma^*(b_n)]$;*

3. *for any individual constant $(a^j)^{T(P)} = (a^j)^{T(P^{A*})}$*

Proof. By induction. ■

Through abduction, this is going to generate a set Δ s.t φ^A is true in some minimal model of $\mathscr{P} \cup \Delta$ and the minimal model of $P \cup \Delta$ satisfies IC. As a result of this, Δ is added to P^A (or KB^A). Since the previous definitions heavily rely on integrity constraints, within the next section I will sketch the intuition behind them. The reader unfamiliar with integrity constraints and their usage in Abductive Logic Programming will find a clear introduction to their usage in appendix A, section A.4.

4.4 Subjunctive conditionals, formally

When we face some unexplained observations, an effective strategy to explain them is to generate plausible hypotheses of the observations through backward reasoning. This can take the form of integrating general rules or specific facts (see Kowalski [65]). The latter can be understood as introducing plausible causes for the relevant observation. The reasoning processes underling abductive reasoning have been formalised in terms of integrity constraints, which capture the idea that an unexplained fact φ in a program P can be explained by multiple sets of clauses $\Delta_1, \ldots, \Delta_n$. The possible *explanans* for φ are in principle obtained by looking at the resolution-tree for φ, and "updating" the program P with the open leaves. This procedure, in general, generates only *relevant* explanations, but it can produce multiple Δs. A choice between alternative explanations can be achieved either with additional constraints on P, or through the cooperation of backward and forward (i.e. deductive) reasoning.

Integrity constraints capture the intuition behind Closed World Reasoning about rules, furthermore their "strength" differs from normal clauses: since the program P has to be updated as to make true the *explanandum* φ, integrity constraints cannot be falsified by a program. The program will be forced to satisfy what the integrity constraint requires.

Within this chapter, I won't dwell much more upon the technicalities of integrity constraints and Abductive Logic Programming (ALP). Henceforth, the formal details of the ALP system used in Kowalski [65] will be freely employed. The reader unfamiliar with integrity constraints and the ALP systems will find an introduction to them in appendix A (see sec. A.4).

I will now proceed to display how the approach proposed for imagination can work for subtractive reasoning. In doing this, I will focus on specific examples taken from "counterfactual" tasks. The choice of analysing these – instead of the spontaneous generation of subjunctive conditionals – on one hand is justified by the abundance of precise data about children performance on these tasks. On the other hand, "counterfactual" tasks, where a choice between possible consequences is offered to children, already suffices to display that Nichols and Stich characterisation of pretend play and imagination as a form of counterfactual imagination cannot be correct, when the latter is understood à la Perner et al. [88] or as subtractive reasoning. The arguments for this claim have already been discussed in chapter 3, hence I will not repeat them here. However, it is relevant to mention that the analysis of spontaneous generation of subjunctive conditionals – which for lack of space and time cannot be fully studied in the present thesis – would represent an extremely interesting possible extension of the approach here defended.

Let me start by introducing an example taken from the literature on subtractive reasoning, which incidentally is also an example of contrary-to-fact reasoning. The examples of "counterfactual" tasks will indeed constitute the motivating examples, and I will use them as the yardstick against which our formalisation needs to be evaluated. The following story is taken from Riggs et al. [96], the experiment they carried out presented it both verbally and visually.

> Jenny is in the garden painting a picture. She leaves the picture on the table, and the wind blows it up into a tree.

After familiarisation with the settings, the children participating to the experiment were asked a control question "where is the picture now?", and a test question "If the wind didn't blow Jenny's picture, where would it be now?". As previously mentioned, notice that the "counterfactual" task based on this story – to be found in Riggs et al. [96] – provides a choice between two consequents for the subjunctive conditional. Therefore, I will label it as prompted subtractive reasoning, to highlight its difference from spontaneous subtractive reasoning.

To formally represent this situation we need to introduce some instances of the axioms of the modified event calculus (EC^*), which formalise the principle of inertia for spatial location, such as :
if the painting is on the *table* at time 0 and no affecting event takes place between 0 and 3, then the painting is on the *table* at 3.

The modified axioms of the Event Calculus are not repeated here, but they are used in the derivations of the relevant queries. A detailed exposition of them can be found in appendix A.3. The affecting event is captured by the predicate $Clipped(l,t,f,l',t')$. The program P will be formulated in the modified constraint structure which augments the five-sorted logic used in van Lambalgen and Hamm [115] with a set of spatial locations S. I will use the constants $table, sky, tree$ to represent spatial locations, while $wind, painting$ are fluent-objects.
$P = (1) Initially(painting, table)$.
$(2) Happens(move, table, 2)$.
$(3) Initiates(move, wind, table, 3)$.
$(4) Trajectory(wind, table, 3, painting, sky, 1)$.
$(5) Terminates(move, painting, table, 2) \leftarrow \exists s < 2(Holdsat(painting, table, s)) \wedge \exists r > 2, \exists y \neq table(Holdsat(painting, y, R))$.
The resolution tree for $?-Holdsat(painting, table, 3)$ on the program P' is represented in figure 4.1.

4.4.1 How subtractive reasoning should work

Now, to model the "counterfactual" task in Riggs et al. [96], the first thing to do is to update the program as required by the assumption that "the wind didn't blow the painting on the tree". A first approximation of one possible manner of doing this is to substitute in the program $Happens(move, table, 2)$ with $Happens(move, table, 2) \leftarrow false$. If counterfactual imagination were sufficient to succeed in the "counterfactual" task, then whichever circumstances making true $Happens(move, table, 2) \leftarrow false$ would work. It is easy to check, however, that this wouldn't lead to any information about where the painting would be.

Since the "counterfactual" task asks the child "where would the painting be?", in order to check as closely as possible how the program P^A s.t. $Happens(move, table, 2)$ has been replaced by $Happens(move, table, 2) \leftarrow false$ would behave, the query $?Holdsat(painting, l, 3)$ is asked, where l is a variable for location. As displayed in fig. 4.4.1 the only computed substitution in a finite branch is $l = table$. Once again, the long resolution for the sub-query $?\neg\exists r < s(Holdsat(f, l, r)) \rightarrow ?\exists r < s(Holdsat(f, l, r)$ FF is left as an exercise to the reader. Since

Figure 4.1: Resolution tree for $Holdsat(painting, tree, 3)$

:- f=painting, l=tree, t=3, s=3, s<t

¬(Clipped(l,s,f,l,t)
r<t, $\neg\exists s < r$, Holdsat(f,l,r),
f= painting,l=tree,t=3, Holdsat(f,l,r), r<t,

[(2)]

$\neg\exists l' \neq l$ Clipped(l,r,f,l',t')
Holdsat(f,l,r), r<t, ¬∃ s<r (Holdsat(f,l,s))
:- f=painting,l=tree,t=3

unsat

Initially(f,l)
f=painting,l=tree,t =0

[Ax.7]

[Ax. 6]

?- (Holdsat(f,l,t)
f=painting,l=tree,t=3

[Ax. 9]

[Ax. 7]

f=painting,l=tree,t=3, t'<t, t=t'+d, l'<l, l=l'+g,
Happens(e,l',t') ,Initiates(e,f_1,l',t'),
Trajectory(f_1,l',t',f,g,d), ¬ Clipped(l',t',f_1,l,t)

[(2)]

f=painting,l=tree,t=3, t'<t, t=t'+d, l'<l, l=l'+g,
e=move,l'=table,t'=2,Initiates(e,f_1,l',t'),
Trajectory(f_1,l',t',f,g,d), ¬ Clipped(l',t',f_1,l,t)

[(3), (4)]

f=painting,l=tree,t=3,t'<t,t=t'+d, l'<l,l=l'+g
e=move, l'=table,t'=2,g=sky,
d=1,¬ Clipped(l',t',f_1,l,t)

sub-derivation

$\neg Clipped(l', t', f_1, l, t) \to Clipped(l', t', f_1, l, t)$ *fails finitely*

:- f=painting,l=tree,t=3,s<t,
Happens(e,l,s), Initiates(e,f,l,t),
$\neg\exists l' \neq l$ (Clipped(l,s,f,l',t)

[(2)]

f=painting, l=tree, t=3, s < t,
Happens(move,l,s), l=table,s=2

unsat

there is no time before 0 s.t. $Holdsat(painting, table, \_)$ holds, it is totally obvious that the query $?Holdsat(f, l, r), r < s, s = 0$ will fail finitely.

After having provided an example of how it is possible to model subtractive reasoning using Constraint Logic Programming, the Event Calculus and Integrity Constraints, I will provide a general definition of how subtractive reasoning is carried out. Therefore, I will check that the definitions proposed are able to display why children between the age of 2 and 5 fail in the more complex subtractive reasoning tasks devised by Perner et al. [88], Cristi-Vargas et al. [20].

Definition 4.11. (Subtractive reasoning) Let φ^A be a consequence of P, and $\varphi^A := \varphi \leftarrow false$. The imagination program P^A is originated from P through σ^* as follows: for any $\chi \in P$

$$\sigma^+(\chi) = \begin{cases} nothing_\chi \in P^A & \text{if } \chi_{[\varphi\rightarrow\varphi/\varphi]} = \varphi \text{ or if } \chi_{\varphi\rightarrow\varphi/\varphi} \vDash_3 \bot; \\ \chi_{[\varphi\leftarrow false/\varphi]} \in P^A & \text{otherwise} \end{cases}$$

$$\sigma^*(\chi) = \begin{cases} \sigma(\chi_{[\psi_i^A/\psi]} \in P^A & \text{if } \psi_i^A \text{ is obtained from abduction over } ?\varphi^A \\ & \text{and } \psi^A \text{ is the result of an alteration of a} \psi \in P; \\ \sigma^+(\chi) & \text{for any } \chi \in P \text{ otherwise.} \end{cases}$$

It is easy to see that the program P^{A*} obtained through σ^* is exactly the one previously described:

$P^{A*} = (1) Initially(painting, table)$.
$(2) Happens(move, table, 2) \leftarrow false$.
$(3) Initiates(move, wind, table, 3)$.
$(4) Trajectory(wind, table, 3, painting, sky, 1)$.
$(5) Terminates(move, painting, table, 2) \leftarrow \exists s < 2(Holdsat(painting, table, s)) \wedge \exists r > 2, \exists y \neq table(Holdsat(painting, y, r)).$

4.4.2 How subtractive reasoning probably works

In the previous section, I have proposed a first sketch of how subtractive reasoning could work in the Jenny's painting example. However, the resolution tree for the simple query $?Holdsat(painting, tree, 3)$ opens an infinite derivation, i.e., the branch opened by ax. 7 in fig. 4.1 leads into a loop. On these basis, I will now propose a second formalisation what the reasoning patters required for providing the "correct" answer to the Jenny's painting story.

> Jenny is in the garden painting a picture. She leaves the picture on the table, and the wind blows it up into a tree. The control question asked was "where is the picture now?", while the test question was "If the wind didn't blow Jenny's picture, where would it be now?"

Initially, I proposed to simply assume that the child understands the test question as asking "where would be painting be,if it was not moved from the table to the tree?". This seemed

Figure 4.2: Where would be painting be, if the wind didn't blow it on the tree?

?- Holdsat(f,l,t),
f=painting, t =3

[ax. 7]

:- t=3, f= painting, Holdsat(f,l,s),
s< t, $\neg\exists$ r<s (Holdsat(f,l,r)), $\neg\exists l' \neq l$(Clipped(l,s,f,l',t)

[ax. 6, (1)]

:- t=3,f=painting, l=table,s=0,
s<t, $\neg\exists$r<s (Holdsat(f,l,r)), $\neg\exists l' \neq l$(Clipped(l,s,f,l',t)

sub-derivation,omitted

:-t=3,f=painting, l=table,s=0
s<t,r<s, $r \neq 1, r \neq 1, r \neq 2 \ldots$
, $\neg\exists l' \neq l$(Clipped(l,s,f,l',t)

sub-derivation

:-t=3,f=painting, l=table,s=0
s<t,r<s, $r \neq 1, r \neq 1, r \neq 2 \ldots$

?- $\neg\exists l' \neq l$(Clipped(l,s,f,l',t) $\rightarrow$?- Clipped(l,s,f,l',t) FF
?-Clipped(l,s,f,l',t), t= table, f= painting, s=0, t=3

[ax. 10]

:- t= table, f= painting, s=0, t=3,Happens(e,g,q),
s<q<t, $l \leq g \leq l'$, Terminates(e,f,g,q) ; Releases(e,f,g,q)

[(2)

:- t= table, f= painting, s=0, t=3
s<q<t, $l \leq g \leq l'$, Happens(move,l,q), $q = 2$

[(2)]

:- t= table, f= painting, s=0, t=3
s<q<t, $l \leq l \leq l'$, false

plausible, on the basis of the fact that the control question was designed to test that the child knows that the painting is on the tree. However, it could also simply be that the child understands the counterfactual supposition "the wind didn't blow Jenny's picture" as a request to update the program in such a way that $?\neg Clipped(table, 0, painting, l', 3)$, where l' is a variable for locations. If this hypothesis is correct, then the child would modify the program carrying out an empty substitution (because there is no occurrence of $Clipped(table, 0, painting, l', 3)$). This would lead to through abduction to the set $\Delta = \{Happens(move, table, 2) \leftarrow false\}$ and hence to the same result as in 4.4.1. However, in this case on the basis of the abduction the following integrity constraint would be added to P^A : IF $?\neg(Happens(move, table, 2))$ *succeeds* THEN $\neg\exists l' \neq tableClipped(table, 0, painting, l', 3)$ *succeeds*. This can be seen in figure 4.3, where the number of clauses used to derive sub-goals refer to the original program P.

On the basis of the integrity constraint IF IF $?\neg(Happens(move, table, 2))$ *succeeds* THEN $\neg\exists l' \neq tableClipped(table, 0, painting, l', 3)$ *succeeds* and $\neg(Happens(move, table, 2))$ is added to the program P. Then the query $?Holdsat(painting, l, 3)$ is resolved in a similar manner as in 4.4.1. However, if the same interpretation is applied to one of the tasks proposed in Perner et al. [88], Cristi-Vargas et al. [20], the program P^{A*} would be modified in such a way that every clause $\psi \in P$ s.t. $\chi \rightarrow \varphi$ would be negated.

Let me consider the following story tested in Perner et al. [88]:

> The floor is initially clean (shown to the child). Carol comes home back home, but she doesn't take her dirty shoes off (the floor becomes all dirty).[5] Just after Carol walks in, also Max gets home. He also doesn't remove his dirty shoes and makes the floor all dirty.[6]

After being presented the story both verbally and visually, each child was asked a *now control question* – i.e. "what does the floor looks like now?" –, a *before control question* – i.e. "what did the floor looked like before the children walked over it?", and the "subjunctive past question" "What if Carol had taken her shoes off? Would the floor be dirty or clean?"[7].

An intuitive formalisation of this making use of the Event Calculus is the following program P_1, where I max, $carol$ are formally properties, and $shoes, floor$ are treated as locations. The formalism I employ would easily allow to mimic the way in which the recognition of a child max is achieved on the basis of properties. However, for the sake of simplicity I assume that max and $carol$ have already been recognised.

(1)$Initially(clean, floor)$.
(2)$Holdsat(carol, floor, 2)$.
(3)$Holdsat(dirty, shoes, 2)$.

[5] A version of the task with only one puppet Carol was tested by Harris et al. [45], obtaining that even 3-years old answered 75% of the past subjunctive questions correctly.

[6] See Perner et al. [88].

[7] With the order of "dirty", "clean" balanced over trials.

Figure 4.3: Abduction over $?\neg\exists l' \neq l Clipped(l,0,f,l',3))$

?- $\neg\exists l' \neq l$ Clipped(l,0,f,l',3) $\rightarrow$?- Clipped(l,s,f,l',t) FF
f=painting,s=0,l= table, t=3

?- f=painting,s=0,l= table, t=3, Clipped(l,s,f,l',t)

[ax. 10]

:- f=painting,s=0,l= table, t=3, 0<r<t, l$\leq g \leq l'$
Happens(e,g,r), (Terminates(e,f,g,r); Releases(e,f,g,r))

no subst. possible

Releases(e,f,g,r), Happens(e,g,r)
0<r<t, l$\leq g \leq l'$, e=move, r=2,
:- f=painting,s=0,l= table, t=3, t'<

:- f=painting,s=0,l= table, t=3,
0<r<t, l$\leq g \leq l'$, Happens(e,g,r),
(Terminates(e,f,g,r); Releases(e,f,g,r))

[(2)]

f=painting,s=0,l= table, t=3,
0<r<t, l$\leq g \leq l'$, e=move, r=2,
(Terminates(e,f,g,r); Releases(e,f,g,r))

[(2)]

:- f=painting,s=0,l= table, t=3,
0<r<t, l$\leq g \leq l'$, e=move, r=2,
$\exists t'$ Holdsat(f,l,t'), t'< 2, $\neg\exists t'' < t'$(Holdsat(f,l,t")

[ax. 6, (1)]

:- f=painting,s=0,l= table, t=3
0<r<t, l$\leq g \leq l'$,e=move, r=2
Holdsat(f,l,0), t'= 0, $\neg\exists t'' < t'$(Holdsat(f,l,t")
Happens(e,g,r)

sub-derivation, omitted

:- f=painting,s=0,l= table, t=3,
0<r<t, l$\leq g \leq l'$,e=move, r=2
t" $\neq 0, t'' \neq 1, t'' \neq 2, \ldots$

:-f=painting,s=0,l= table, t=3,
0<r<t, l$\leq g \leq l'$, Happens(e,g,r)
(Terminates(e,f,g,r); Releases(e,f,g,r))

[(5)]

:- f=painting,s=0,l= table, t=3,
0<r<t, l$\leq g \leq l'$,e=move, r=2
Happens(e,g,r), $\exists t'$ Holdsat(f,l,t'), t'< 2, $\neg\exists t'' < t'$(Holdsat(f,l,t")

[ax. 6, (1)]

:- f=painting,s=0,l= table, t=3
0<r<t, l$\leq g \leq l'$,e=move, r=2
Holdsat(f,l,0), t'= 0, $\neg\exists t'' < t'$(Holdsat(f,l,t")
Happens(e,g,r)

sub-derivation, omitted

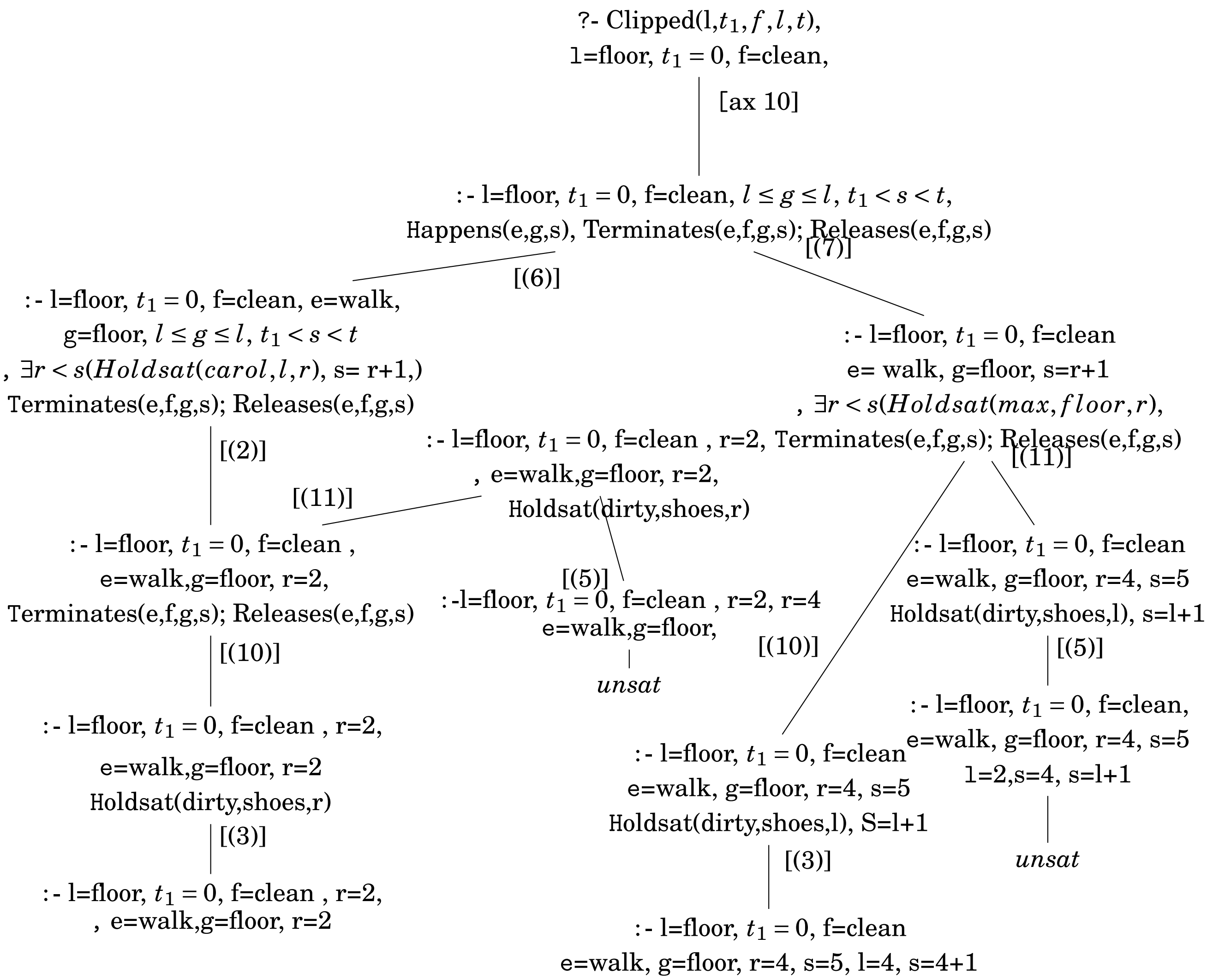
:- f=painting,s=0,l= table, t=3,
0<r<t, l$\leq g \leq l'$,e=move, r=2
t" $\neq 0, t'' \neq 1, t'' \neq 2, \ldots$
Happens(e,g,r)

[(2)]

:- e= moved, g= table, r=2

Figure 4.4: Abduction over ?$\neg\exists\exists t < 7(Clipped(floor, 0, clean, floor, 7))$

?-$\neg\exists t < 7(Clipped(floor, 0, clean, floor, 7))$ → ?- (Clipped(floor,0,clean,floor,t)) FF



(4)$Holdsat(max, floor, 4)$.

(5)$Holdsat(dirty, shoes, 4)$.

(6)$Happens(walk, floor, t) \leftarrow \exists s < t(Holdsat(carol, floor, s) \wedge t = s+1)$.

(7)$Happens(walk, floor, t) \leftarrow \exists s < t(Holdsat(max, floor, s) \wedge t = s+1)$.

(8)$Initiates(walk, dirty, floor, t) \leftarrow \exists s < t(Happens(walk, floor, s) \wedge Holdsat(dirty, shoes, s) \wedge t = s+1)$.

(9)$Initiates(walk, dirty, floor, t) \leftarrow \exists s < t(Happens(walk, floor, s) \wedge Holdsat(dirty, shoes, s))$.

(10)$Terminates(walk, clean, floor, g) \leftarrow \exists r < g(Holdsat(dirty, shoes, r) \wedge g = r+1)$.

(11)$Terminates(walk, clean, floor, g) \leftarrow \exists r < g(Holdsat(dirty, shoes, r) \wedge g = r+1)$.

Now, in the case of $Carol$: from (2),(6) we obtain that $Happens(walk, floor, 3)$. Furthermore, from this result, (3),(8) it is possible to derive that $Initiates(walk, dirty, floor, 4)$. Finally, from axiom 10 and (3), it is obtained that $Clipped(floor, 0, clean, floor, 4)$ and hence that $\neg Holdsat(clean, floor, 5)$. A similar result can be obtained for the case of *Max*. Now if the child understands the question “what if Carol had taken off her shoes” not as $Holdsat(dirty, carol, 2) \leftarrow false$ but as $\neg\exists t' < 7(Clipped(floor, 0, clean, floor, t))$, then the substitution executed and abduction entail that she believes that both Carol and Max didn’t make the floor dirty with their shoes. The reasoning process behind the settings of this process of reasoning towards an interpretation is then reasonably formalised by the following steps:

1. Assume that $\neg\exists t' < 7(Clipped(floor, 0, clean, floor, t))$ *fails*.
2. Carry out abduction over the query $?\neg\exists t' < 7(Clipped(floor, 0, clean, floor, t)) \rightarrow \exists t' < 7(Clipped(floor, 0, clean, floor, t)$ *FF*. The set obtained is $\Delta = \{Holdsat(dirty, carol, 2) \leftarrow false; Holdsat(dirty, max, 4) \leftarrow false\}$;
3. the program so obtained ${P_1}^{A*}$ can be simplified by adding the IC: $\exists t' < 7(Clipped(floor, 0, clean, floor, t) \leftarrow false$.

The latter extension simplifies noteworthy the complexity of a derivation of $?Holdsat(clean, floor, 7)$. Figure 4.4 displays the abductive derivation over $?\neg\exists t' < 7(Clipped(floor, 0, clean, floor, t))$, which gives as a result $\neg(Holdsat(dirty, carol, 2))$ and $\neg(Holdsat(dirty, max, 4))$. This means that $Holdsat(dirty, carol, 2) \leftarrow false$ and that $Holdsat(dirty, max, 4) \leftarrow false$. As a side note, notice that because this reasoning process ensures that $\exists t' < 7(Clipped(floor, 0, clean, floor, t)$ fails, then it would be reasonable to use this reasoning process to simplify the program and to simply substitute the integrity constraint $\neg(Holdsat(dirty, carol, 2)) \wedge \neg(Holdsat(dirty, max, 4)) \rightarrow \neg\exists t' < 7(Clipped(floor, 0, clean, floor, t))$ with $\exists t' < 7(Clipped(floor, 0, clean, floor, t)) \rightarrow false$.

4.5 Pretend Play, formally

In the previous sections, I have introduced the general formalism necessary to model the human ability to imagine alternatives and to construct scenarios for these imagined clauses. This general framework will now be used to tackle the formalisation of pretend play. As it has been hinted in chapter 3, one of the grounding hypotheses of the current thesis is that closed world reasoning about rules and closed world reasoning about facts can explain a relevant part of the difference in pretence behaviour between neurotypical and autistic children. More specifically, I will formally distinguish spontaneous pretend play from prompted pretend play on the basis of the distinction between CWRr and CWRf. This approach is founded on the idea that in elicited or prompted pretend play the child is explicitly “forced” or required to construct a pretend model of the utterances (alternatively, actions) of the experimenter, caretaker or playmate. This can be understood in the following terms: the children when prompted to engage in a specific form of pretend play is asked to assumed that the pretence premise is true.

Before starting off with the formal work, it is important to stress once more the differences among different types of make-believe games: overlooking these distinctions would lead me to conflate the processes underlying these diverse kinds of pretending. Just to give a flavour to the importance of conceptual precision, consider what would happen if I equated individual, social and elicited pretending. In individual pretend play – as it has been defined in section 3.5 – the child is freely developing the imagined alternative and it is creating the imaginary circumstances for it without the explicit intervention of someone else. In social pretend play – and even more in prompted conditions –, the child is implicitly given the imaginary alternative and some cues about its circumstances. It is extremely revealing that children often spend various minutes discussing on the acceptable settings for their play, before actually starting engaging in pretend actions and behaviour. Finally, if the child is elicited by an experimenter or caretaker to engage in a specific kind of pretend behaviour, the child is forced to assume that the pretence premise is true. I hypothesise that the mechanisms in play in elicited pretence within neurotypical children and the autistic population diverge in that autists seem to treat the pretence premise as a premise, i.e. as a statement that they are asked to assume even though it doesn't make much sense for them. Neuro-typical children, on the other hand, are prompted to engage in the same kind of creativity that is employed normally in storytelling and make-believe plays, just on more complex levels such as object substitution and more generally symbolic play.

By conflating these forms of pretend, I would be assuming that there is no relevant cognitive mechanism between imagining something and behaving along its guide and processing non-literal utterances of someone else and make sense of them. Even more, I would be implying that eliciting to pretend more have no effect whatsoever on children[8].

The background theory I have been hinting at in 3.5.3 and in A.4 is formally represented by the Event Calculus (with the extended axioms EC^*) insofar as causality in general is concerned, and scenarios when specific micro-theories are concerned. Formally, I represent scenarios as part of the knowledge base (KB). The distinction between knowledge base and set of clauses indeed serves to ease the presentation of the distinction between facts and rules. In any case, drawing a line between a knowledge base and a set of atomic facts may have some cognitive plausibility, and one could further refine the distinction by differentiating scenarios as coming from episodic memory – and hence representing scripts – from the general rules of causality that would be generated inductively. Within the present writing, the scenarios have been directly constructed from the examples considered. As a future possible extension of the work, one could use the *Cyc*[9] knowledge base to check what a big *corpora* would suggest for many situations, and how this

[8] And this is obviously false, as it can be seen from the following studies: Kavanaugh and Harris [62],Harris [43], Scott et al. [104],Jarrold et al. [58],Jarrold [56],Jarrold and Conn [57],Jarrols et al. [59],Peterson and D.M.Bowler [89],Bigham [9] *et cetera*.

[9] See http://www.cyc.com/kb/

would relate with the axioms of the Event Calculus.

4.5.1 Behaviour: pretence production, formally

Pretence production is formally more difficult than pretence understanding to represent, because the ability to produce pretended acts requires the ability to understand the pretence stipulation(s) and to handle various components at the same time. As I attempted to argue in chapter 3.5.4, I am convinced that to define an act of play as pretence it is necessary to both be able to relate the make-believe representations to the “reality file”, and to represent the pretence as an event limited in time. The first requirement does not entail that the child has to be (theoretically) able to distinguish reality from pretence, but rather that the child must

1. be able to unfold a causal chain;
2. be able to realise that pretend actions have different effects from real actions;
3. be able to effectively produce these actions.

On the basis of this conceptual definition, I will treat a conscious distinction of reality and appearance as a superfluous condition. Requiring children to conceptually distinguish what is “real” from what is “pretended” or “fantasy” doesn’t take into account children’s usage of the words “real”, “really” and “pretend”[10].

In order to display how the pretend program is related to the reality program, and how children are able to behave appropriately to their pretence stipulations, a new predicate $Simulate$ is added to the language. Intuitively, the predicate $Simulate$ imports the pretence program P^{A*} and allows the person engaging in the act of imagination to act accordingly to her pretend stipulations.

Definition 4.12. (Simulation) Through Gödel coding, the pretence program P^{A*} is translated into a finite set of atomic formulas. Take the conjunction of this finite set and call it $\bigwedge_{\varphi^A \in P^{A*}} \ulcorner \varphi^A \urcorner = P'$. Because nominalisation retains some temporal information about $\ulcorner \varphi^A \urcorner$ and all formulas in the language are temporally index, extract the first time t occurring in the pretend program and the last time s occurring in the pretend program. Then $Simulate$ has as objects:

$$Simulate(P', t, s).$$

and $Simulate(P', t, s) \in \mathscr{P}$.

On the basis of the predicate $Simulate$, the pretend premise can be defined.

Definition 4.13. (Pretence Stipulation) A pretence stipulation or premise is a formula $Holdsat({a_i}^A, l, t)$ where ${a_i}^A$ is an alteration to a b_i in original program P as defined in def. 4.3, l is a spatial

[10] Luckily there are some studies on this topic, such as Bunce and Harris [11] and Harris et al. [47].

location, and t is the first time occurring in P^{A*} s.t. $[{a_i}^A/b_i]$. The pretence stipulation is defined as follows: $Holdsat({a_i}^A), T) \leftarrow Simulate(P', T, D) \wedge Holdsat(\chi(b_i), l, t$ for $\chi({a_i}^A, T) \in P'$.

This definition mirrors the informal intuition that when engaging in a form of pretend play, on the basis of a verbal/non-verbal, explicit or explicit that starts off the pretence episode, some substitution over (i.e. identification between) fluent-objects is established for a certain interval of time. Obviously there will be many pretence premises that will be derived from the pretence program. These definitions entail that all the substitutions $[\psi^A/\psi]$ carried out in P^{A*} are imported in the original program P in the form of ${a_i}^A$ where $\ulcorner \psi^A \urcorner = {a_i}^A$. Furthermore, in the case of spontaneous pretence it is easy to see that $Holdsat({a_i}^A, l, t)$ will be problematic to integrate.Intuitively, the present formalisation suggests that a pretence premise establishes that for a limited amount of time b is ${a_i}^A$ as long as this doesn't create contradictions or inverts the substitution between ${a_i}^A$ and b. Instead, in the case of elicited pretence some statement $Holdsat(\psi({a_i}^A), T)$ is forced upon the subject by an integrity constraint of the form $Holdsat(\psi({a_i}^A), l, t)$ succeeds .

Definition 4.14. Pretence behaviour is defined as follows:

$$(1) \forall \varphi^A \in P' \text{ s.t. } \varphi^A = Initiates(e, f, l, t) \text{ or } \varphi^A = Happens(e, l, t)$$

$$\text{or } Terminates(e, f, l, t) \text{ or } Releases(e, f, l, t) \text{, then } \varphi^A \in \mathscr{P}^{11}$$

The end of the pretence episode is defined as :

If φ^A is a pretence premise, according to def. 4.13, and S is the last time in $Stipulate(P', T, S)$, then $Terminates(empty\_action, \varphi^A, S)$.

This definition informally means that all the events initiating or terminating a time-dependent property are imported in the program $\mathscr{P}$. This also means that all the actions required to produce the appropriate pretence behaviour become *real*, and can be executed.

So far I have introduced the general approach to model pretend play, however, as I have tried to display previously, the differences between different forms of pretend play cannot be left undetermined in a formalisation of it, without losing much grasp of the phenomena under consideration. Therefore, I will now show how the general method and the definitions proposed in section 4.2 allow to formalise object substitution for both neuro-typical and autistic children. Notice that, according to the definitions advanced in section 4.2 and to the dissociation hypothesis, object substitution will be generally the most requiring type of pretence for children with autism. This can be easily read off from the definitions, since only object substitution is always defined in terms of an incompatibility between a fact about the world $\varphi(a)$ and a pretence stipulation

[11] This condition can be understood as set of integrity constraints of the form $?Stipulate(Initiates[e, f, l, t], l', t')$ succeeds , then $Initiates(e, f, l', t')$ succeeds, and hence it can be forced upon $\mathscr{P}$.

$\varphi(b)$ which cannot be simultaneously true. This is not the case in many instances of attribution of absent properties or imagination of absent entities, for they require the reality representation to be augmented with a fact that doesn't necessarily contradict what the child previously knew (and still does) about the world.

Notice that, following the categorisation of object substitution as one of the core examples of symbolic play, most of the literature concerning pretend play agrees that object substitution constitutes a rather complex instance of pretending, and that it is particularly hard for younger children and children with autism to produce the appropriate responses.

4.5.2 Object substitution

According to the definition previously proposed of object substitution – i.e. def. 4.5 –, a fluent-object is replaced by another fluent object. This entails that, following the definition of pretence stipulation (def. 4.13), the child establishes a certain replacement of a fluent-object a with another fluent object b that holds for a certain interval of time, that is until the pretend play episode ends. Let me focus on one of the most paradigmatic examples of object substitution: a child pretends that the banana in front of her is a phone and entertains a (pretend) phone conversation through it, before hanging up and hereby bringing the pretend episode to a close.

Informally, within the present framework, one would expect children to substitute all the known facts about the banana, e.g. that it is laying on the table in front of me, that it is a banana, that it is yellow *et cetera*, with the corresponding statements about a phone unless the substitution leads to a contradiction or would allow to infer back that the object is a banana. For instance, I expect children to behave as if the object is a phone, and the phone occupies the same location the banana were occupying, but to ignore the fact that the banana is yellow and that has a banana-shape, because one might have a rule according to which something yellow and with a banana-shape is a banana.

With the awareness that it is obviously a theoretical and *post hoc* distinction, and with all probabilities children's reasoning patterns are much more continuous, for the sake of clarity and to keep the formal explanation neat, I will distinguish four steps in the pretend production. This will help in identifying which steps are problematic – according to the dissociation hypothesis – for children with autism. This will also allow me to focus only on the problematic cases in the autism's case.

Let me start by introducing the knowledge base, which will represent the scenario and the general rules about causality that will be used.

$KB :=$ {Axioms of the EC^*,

(1) $\forall l, \forall t(Holdsat(piece\_food, l, t) \leftarrow (Holdsat(banana, l, t)))$.

(2) $\forall l, \forall t(Holdsat(banana, l, t) \leftarrow Holdsat(smooth, l, t) \wedge Holdsat(prolate, l, t) \wedge Holdsat(l\_shape, l, t)))$.

(3) $\forall l, \forall t(Holdsat(phone, l, t) \leftarrow Holdsat(prolate, l, t) \wedge Holdsat(l\_shape, l, t) \wedge Holdsat(keys, l, t))$.

(4) $\forall l, \forall t(Happens(call[A], l, t) \leftarrow \exists s < t(Happens(take[A], l, s)) \wedge \exists t' \leq t(Happens(talk[A], l, t') \wedge$

$s < t') \wedge Holdsat(phone, l, t))$.
(5) $\forall l, \forall t(Happens(eat[A], l, t) \leftarrow \exists s < t(Holdsat(piece\_food, l, s) \wedge \neg\exists r < s(Holdsat(piece\_food, l, r) \wedge$
$\neg\exists l'(Clipped(l, t, piece\_food, l', t')))))$.
(6) $\forall l, \forall t(Happens(invite[A, I], l, s) \leftarrow Happens(call[A], l, s) \wedge (A \neq I))$.
(7) $\forall l, \forall s(Happens(call[I], l, s) \leftarrow \exists A \exists t < s(Happens(invite[A, I], l, t)))$.
(8) $\forall l, \forall t(Terminates(call[A], banana, l, t) \leftarrow Holdsat(phone, l, t))$.
(9) $\forall l, \forall t(Terminates(call[A], t, smooth, l, t) \leftarrow (l_1 \leq l \wedge l_2 \leq l \wedge Happens(call[A], l, t)))$.}

The clauses (1)–(3) are formally definitions of fluents[12]. Furthermore, notice that clauses (2) and (3) characterise the recognition of an object on the basis of features co-located spatio-temporally. For example, the object banana is recognised on the table only if its prototypical features are processed as being on the table and bound together into an unitary representation, which can also be thought of as a conjunction. Furthermore, as the set of clauses P displays, the object banana changes its position if its features do.

Let me know introduce the set of facts.
$P := \{(1^*) Initially(smooth, l)$.
$(2^*) Initially(prolate, l)$.
$(3^*) Initially(l\_shape, l)$.
$(4^*) Happens(take[mum], l, 1)$.
$(5^*) Happens(talk[mum], l, 3)$.
$(6^*) Initiates(take[mum], move, l, 1)$.
$(7^*) \forall f, \forall l, \forall t, \forall g(Trajectory(move, l, t, f, g, 1) \leftarrow (f = l\_shape) \wedge l + 1 = t' \wedge l + g = l')$.
$(8^*) \forall f, \forall l, \forall t, \forall g(Trajectory(move, l, t, f, g, 1) \leftarrow (f = smooth) \wedge t + 1 = t' \wedge l + g = l')$.
$(9^*) \forall f, \forall l, \forall t, \forall g(Trajectory(move, l, t, f, g, 1) \leftarrow (f = prolate) \wedge t + 1 = t' \wedge l + g = l')$.
$(10^*) \forall t, \forall l(Terminates(take[mum], smooth, l, t) \leftarrow \exists l' \neq l, \exists t' > t(Holdsat(smooth, l', t'))$.
$(11^*) \forall t, \forall l(Terminates(take[mum], l\_shape, l, t) \leftarrow \exists l' \neq l, \exists t' > t(Holdsat(l\_shape, l', t')))$.
$(12^*) \forall t, \forall l(Terminates(take[mum], prolate, l, 1) \leftarrow \exists l' \neq l, \exists t' > t(Holdsat(prolate, l', t')))$.
$(13^*) Happens(invite[mum, I], l', 4).\}$

Let me start displaying what the program $\mathscr{P} = P \cup KB$ means. For the sake of brevity, I will do so relying on the completion of the program (see def. A.38), which I denote by $comp(\mathscr{P})$, and of a theory $\mathscr{T}$, whose details are introduced in app. A. The completion of the program makes true that the object in front of the child at time 0 is a banana, i.e. $\mathscr{T} + comp(\mathscr{P}) \vDash_3 Holdsat(banana, l, 0)$. Moreover, at time 2, as a result of the actions of the mother picking up the banana and bringing it to her, the banana changes position– or to be extremely precise the features, which allow

[12]Notice that $keys, smooth, prolate$ etc are fluents. This means, formally that they are functions from time to truth values. Let sim be a function defining the situation of an object, i.e. which gives either the object as a fluent or its defining feature. For example, $(x, l, t) \mapsto \forall t(x, t)$, where t is a temporal location. Through abstraction, the function becomes able to be the object of a defined predicate, and its spatio-temporal parameters can be retained into the formula.

the child to recognise a banana, change position, and as soon as this action is completed the banana is recognised again. This means that $\mathscr{T} + comp(\mathscr{P}) \vDash_3 Holdsat(banana, l', 3)$ and that $\mathscr{T} + comp(\mathscr{P}) \nvDash_3 Holdsat(banana, l, 3)$.

It is now possible to start with the exposition of the four steps involved in the banana-phone play. As it will be remarked as we proceed, every step corresponds to one of the four elements of pretence defined in the previous parts of the chapter, i.e. σ, σ^+,σ^* and the import from the pretence file through $Simulate$ (see def. 4.14).

Step 1: As it can be read off from P, I am going to model the banana-phone example, under the assumption that the mother starts the play and the child must understand what is going on. The child sees her mother picking up the banana and talking into it.

$$(i)\ Happens(take[mum], l, 1).$$

$$(ii)\ Happens(talk[mum], l', 3)$$

On the basis of (i),(ii) and $(7^*) - -(9^*)$, is is derived that the object-banana moves with its features, and hence that

$$(iii)\ Holdsat(banana, l', 3)$$

Furthermore, at time 4, the child is invited to "join" the phone conversation, i.e. $\mathscr{T} + comp(\mathscr{P}) \vDash_3 Happens(invite[mum, I], l', 4)$. Notice that on the basis of clause (6), this entails that

$$(iv)\ Happens(call[mum], l', r), r < 4.$$

In order to understand what is going on, the child does abduction over $?Happens(call[A], l', 3)$, and she obtains that her mother cannot be having a phone call because the banana is smooth, while a phone has keys.

At this point, the neuro-typical children inhibits the attentional features –i.e. the smoothness of the banana– on the basis of its use. This is modelled via (9): on the basis of

$$(iv) Happens(calling[mum, l', 3),$$
$$Holdsat(smooth, l', 3)$$

Axiom 10 $Happens(e, g, s) \wedge t < s < t' \wedge l \leq g \leq l' \wedge (Terminates(e, f, g, s) \vee Releases(e, f, g, s)) \rightarrow Clipped(l, t, f, l', t')$

the child infers that $\neg Holdsat(smooth, l', 3)$ and she is able to substitute $Holdsat(smooth, l', 3)$ with $Holdsat(keys, l', 3)$. Notice, *en passant*, that this is an allowed substitution according to the definition 4.3.

Step 2: Once the substitution of $[keys/smooth]$ has been carried out within $Holdsat(smooth, l', 3)$. The first step to engage in the pretence production is to check where it is possible to propagate the substitution within P, without generating contradictions nor deriving back that $Holdsat(smooth, l', 3)$. This generates the following clauses, which replace $(1^*), (8^*), (10^*)$

$(1^{*A}) Initially(keys, l)$.
$(8^{*A}) \forall f, \forall l, \forall t, \forall g(Trajectory(move, l, t, f, g, 1) \leftarrow (f = keys) \wedge t+1 = t' \wedge l + g = l')$.
$(12^{*A}) \forall t, \forall l(Terminates(take[mum], keys, l, 1) \leftarrow \exists l' \neq l, \exists t' > t(Holdsat(keys, l', t')))$.
Now, the child is also able to infer that $Holdsat(phone, l', 3)$.

Step 3: At this point, the child may further enlarge the minimal circumstances for the pretend play by doing abduction. In this specific case, abduction is not strictly needed. Once the child has that $Happens(call[mum], l', 3)$ and that $Happens(invite[mum, I], l', 4)$, she can infer from (9) that

$$(v)\ Happens(call[I], l', 5)$$

Step 4: On the basis of def. 4.14, the child can now import the action $Happens(call[I], l', 5)$.

Formally, all the formulas in P^A of the form $Happens(e, l, t)$ or $Initiates(e, f, l, t)$, $Terminates(e, f, l, t)$, $Releases(e, f, l, t)$ should be imported within $\mathscr{P}$. However, because all other formulas except $Happens(call[I], l', 5)$ are already in P, the only effective addition – in accordance with def. 4.14 – is $Happens(call[I], l', 5)$.

The steps 1–4 display how the relevant acts can be produced. Furthermore, on the basis of the definition of 4.13, the child can infer the tags or pretence stipulations that allow her to answer to questions about what happened in the pretend episode. Finally, what terminates the pretence stipulation is not defined by $\mathscr{P}$, on the basis of the intuition that the child keeps constructing the pretended scenario as long as she is able to. In accordance with def. 4.14 when the child is not able anymore to enlarge the pretend program with novel actions and events, the pretend play finds its natural close[13]. The pretence stipulations however don't hold true, and hence they do not propagate over future events and actions the child engages in.

4.6 Object substitution: the autistic case

Within this section, I will attempt to display what changes, according to the formalisation proposed, when a children with autism is involved in an object substitution. It should be noticed, once more, that object substitution is formally the most requiring type of pretend play, because it requires the child to relate two conflicting facts about the world.

Let the program P and KB be as in section 4.5.2, except from clause (9) which is not in KB, and clause (5) which is modified as follows:
$(7')(11)\ \forall l, \forall t(Happens(eat[A], l, t) \leftarrow \exists s < t(Holsat(piece\_food, l, t) \wedge \neg \exists r < s(Holdsat(piece\_food, l, r))))$.
These changes are motivated by the idea that children with autism have difficulties with handling abnormalities. Clause (9) represents an exception to the general rule, according to which the function of an object is dictated by its characteristic attentional features, and in particular by its

[13] This doesn't entail that it is *impossible* to define events and situations that will force the pretend play to end. Doing this is quite straight-forward, but for the analysis of children's ability to spontaneously pretend the case in which the child is left free to pretend as long as she can or wants is more interesting. As mentioned in chapter 2 and section 2.6, various studies display that children with autism spend significant less time in spontaneous pretence.

shape (see Landau et al. [68], Deák et al. [22] and within pretend play Bigham [8]). Similarly, clause (5) represents the general rule according to which if an object is food and its being food is not affected by some event, the agent may eat the food.

I will display that for children with autism, already the first step is problematic, and the difficulties with this first step explain why children with autism seem to struggle in engaging with pretend production.

Step 1: Just as in the neuro-typical case, the child infers $(i)--(iv)$. From (iv), as previously mentioned, the child can decide to do abduction over $Happens(call[mum],l',3)$. Once more, the child infers that her mother cannot be having a phone conversation, because the object-banana is smooth, while a phone is not. The crucial step which allowed neuro-typical children to inhibit the smoothness of the banana, and to substitute $Holdsat(keys,l',3)$ for $Holdsat(smooth,l',3)$ crucically relied on (9). In its absence, the child has that

$$(iv) Happens(calling[mum,l',3),$$
$$(4)\ Holdsat(smooth,L,T) \leftarrow \forall L_1,L_2 \leq$$
$$L(Holdsat(slope,L_1,T) \wedge Holdsat(slope,L_2,T) \wedge \neg(Terminates(L_1,T,slope,L_2,T))) \text{ and}$$
$$\text{Axiom 10 } Happens(e,g,s) \wedge t < s < t' \wedge l \leq g \leq l' \wedge (Terminates(e,f,g,s) \vee Releases(e,f,g,s)) \rightarrow Clipped(l,t,f,l',t')$$

but she is unable to relate them, i.e. no way of inhibiting the fact that the banana is smooth and a phone is not.

Therefore, the children with autism does not manage to switch to a non-literal interpretation, and hence engage in the pretend play. Notice that the difficulty of children with autism to understand pretence in cases where the form of the prop differs from its imagined target is supported by the results in Bigham [8, 9]. Furthermore, various reports on highly function children with autism describe them arguing that the object cannot be a phone, because it is a banana, it has the shape of a banana, it does not have keys *et cetera*[14].

[14] The reader may wonder which is the difference between this case and counterfactual reasoning: in both cases the predicate $Clipped$ works as the abnormality, i.e. what makes an object stop having a certain property. One could think that if $Clipped$ works in one case than it should work in all cases. However, the neuropsychological studies on autism suggest that the condition is accompanied by less inhibitory synapses relative to excitatory ones (see van Lambalgen and Stenning [117, p. 284]). This doesn't mean that inhibitory synapses are completely absent, but rather they are less numerous. Furthermore, according to the hypotheses in Frith [31] and Huttenlocher [53], this would vary among different areas of the brain. For the present thesis this means that the dissociation hypothesis is not an *all-or-nothing* assumption, but that it can vary across tasks. Notice that if people with autism had no inhibitory synapses, it would be impossible for them understand changes of location and, more generally, to carry out structured actions. The assumption that in this specific case (the banana-phone) there is a problem of exception-handling is justified by reports describing children with autism explicitly discussing that the object *cannot* be a phone, because it is a banana. The following is one example of this report from a mother of a child with Asperger's syndrome:

> It used to be that if you gave my son a banana and told him it was a phone he would sit there and stare at it, or perhaps just eat it! Okay, that is an appropriate response; I will give him that. But if I tried to initiate playing, to make-believe that the banana really was a phone, he would argue with me and go into great detail as to how exactly a phone and a banana were completely different. His brain is not wired to think that way, so he couldn't see anything but a banana or a phone. [...]

This displays why children with autism appear to struggle in engaging in object substitutions, and pretend plays which use objects with a different form from their pretended role. Let me briefly hint at how this formalisation relates to the literature on pretend understanding. The tests of pretence understanding carried out by Kavanaugh and Harris [62] and Jarrold et al. [58] appeared to display that autistic children's understanding of pretence is intact. Notice, however, that the kinds of pretence tested in [62] and Jarrold et al. [58] are mostly cases of attribution of absent properties or imagination of absent entities (already noticed in Bigham [8, p. 267]). In accordance with the formalisation I advanced, recent studies (see Bigham [9, 9]) suggest that there is a difference in understanding between object substitution and imagination of absent objects or properties (see Bigham [8, p. 35], table 3).

With regards to production of object substitution, the results in Charman and Baron-Cohen [18],Jarrols et al. [59] and Lewis and Boucher [74] display that the difference in object-substitution production in *prompted* conditions within the autistic group relative to the controls or mentally handicapped is not significant. These results concern the capacity to produce *some* action appropriate to the object substitution, not the novelty of acts generated by the subjects. In the latter case, the difference becomes significant (see Charman and Baron-Cohen [18, p.328]). The analysis of elicited object-substitution will be the topic of the next section.

4.6.1 Elicited object substitution

In the case of elicited or scaffolded object substitution, I hypothesise that the difficulties of children with autism are partially overcome because the pretence stipulation is forced upon the subject by means of an integrity constraint.

To the best of my knowledge, the very few studies on the performance of children with autism in elicited object substitution suggest that elicitation and scaffolding allow children with autism to perform comparably to normally developing children. According to the theory here proposed, I suggest that the difference between spontaneous and elicited pretence has a formal counterpart in the use of integrity constraints. This means that the children with autism can be brought to engage with object substitution – even though this may require very explicit elicitation and quite some time. Furthermore, I suggest that even in the case where the children with autism succeeds in entertaining an alternative (imaginative) representation of the world, the behaviour will suffer from the impairment in CWRf. This explains why the behaviour is more stereotypical, and why children with autism are sometimes reported to repeatedly try to eat a block that is pretended to be food.

More formally, *Step 1* would succeed because of an integrity constraint of the form

$$?Happens(call[mum], l', 3) \text{ succeeds } ?\ Holdsat(phone, l', 3) \text{ succeeds }.$$

This means that the difficulties with the inhibition of the actual form of the object, which children with autism display in spontaneous pretence, are bypassed. Integrity constraints, or better CWRr, are assumed to be intact in children with autism. This means that the child is now able to "open" an alternative representation of reality, i.e. one where the object at l' at 3 is a phone.

The success in accepting the pretence premise, however, does not guarantee that children with autism in elicited pretence will behave as controls.

Let me assume that the steps *2-3* are just as in the case of neuro-typical children. At the fourth step, when the actions generated by the pretend play are imported into the "reality program", the child with autism will import that

$$(vi) Happens(call[I], l', 5).$$

However, at this point, the modification of clause (7) will affect the performance of children with autism. Since clause (5) does not envision abnormal situations, the fact that $Holdsat(phone, l', 5)$ may not be able to inhibit the typical tendency to eat an object if it is food[15]. Suppose that the child imports (vi), in $\mathscr{P}$ she still has that

$$Initially(banana, l),\ Holdsat(banana, l', 5), \text{ and}$$
$$\forall l, \forall t(Happens(eat[A], l, t) \leftarrow \exists s < t(Holsat(piece\_food, l, s) \wedge \neg \exists r <$$
$$s(Holdsat(piece\_food, l, r)))).$$

From which, she may decide to eat the food, i.e. $Happens(eat[I], l', 5)$.

This perseverative behaviour is not always observed in autistic children's pretending. As previously mention, autism is a label for a huge spectrum of conditions, which are obviously going to differ slightly in skills and behaviour. Furthermore, I do not see the *dissociation hypothesis* as all-or-nothing, but rather as a regulative guide to the understanding and explanation of reasoning processes in autism. This means that the level of perseveration observed in pretending will vary on the basis of chronological and verbal age of the children studied.

In any case, I suggest that this example displays how the production of pretended actions can be affected by an impairment in closed world reasoning about facts.

4.6.1.1 Some hypotheses about pretend play in autism

In the characterisation of *step 3* above, for the sake of simplicity, I assumed that the reasoning processes emplyed by children with autism were exactly the same as the ones proposed for neuro-typical children. I am now going to briefly display how also the generation of novel pretence actions may be affected by an impairment in CWRf.

Suppose that we were to extend $\mathscr{P}$ with the following clauses:
(10) $\forall l, \forall s(Terminates(destroy, phone, l, s) \leftarrow Happens(destroy, l, s))$.
(11)$\forall l', \forall t(Terminates(hang\_up[A], l', t) \leftarrow t = 6 \wedge Happens(hangs\_up[A], l', t))$.

[15]Notice that program $\mathscr{P}$ idealises away from reality under many aspects: for instance, it would be more appropriate to eat something if the object is food and the subject is hungry.

$(12)\forall l', \forall t(Initiates(calling[I,X], l', t) \leftarrow t = 7 \wedge \exists s \leq t(Happens(calling[I,X], l', s)))$.

On the basis of (12)–(14) the child now has plenty of possibilities on how to continue with the pretend play. She might decide that her interlocutor hangs up – i.e. adding $Happens(hangs\_up[A], l', 6)$. According to the Axioms of the EC^*, one would obtain that the phone conversation is terminated. This means that the it stops being true that there is a phone conversation, and hence there are no further ways to extend the pretence program. The child may decide to add to P^A the following fact:

$$(vi)\ \ Happens(destroy[I], l', 8)$$

and she knows that this will have an effect not only on the phone conversation but on the fact that the object at 8 located at l' is a phone.

Under the assumption that children with autism have an impaired ability to deal with abnormalities, the child with autism who pretends that her interlocutor decides to hang up may have a simple conditional rule of the form "if the interlocutor hangs up then there is no phone conversation going on". But even if this is not problematic for the child with autism, she might have no way to relate the possibility of destroying the phone with the fact that she is pretending that the obect at l', at time 8, is a phone.

This means, that *contra* the hypothesis in Jarrols et al. [59], children with autism may not impaired in the lack of generativity itself, but this might an apparent effect of the fact that spatio-temporally dependent properties are not easily modified according to their minds. According to the definition of pretence behaviour proposed (def 4.14), it is possible to hypothesise that, because of this difficulty in modifying time-dependent properties, children with autism are more prone to abandon pretend earlier than neuro-typical children. Furthermore, it will be more likely to apply stereotypical scripts and execute them without altering or extending them.

As I have hinted above (see sec. 4.5.1), on the basis of both the definitions proposed in section 4.2 and the consensus within the literature on pretend play, I assume that object substitution constitutes the most difficult form of pretence. As far as the logical form is concerned, this can be easily seen by looking at the definitions 4.5 - 4.7. In most cases imagination of absent properties and imagination of absent entities will be less requiring especially for children with autism, because it will be possible to simply augment their representation of reality without having to detect and inconsistency. Furthermore, as it appeared clearly from the analysis of the banana-phone example, most instances of pretend production involve some pretence comprehension. More specifically, the difficulty identified in section 4.6 for the banana-phone play was actually a problem in the comprehension of the pretence carried out by the mum. This shouldn't come as a surprise, considering that most authors (see Leslie [71], Jarrold et al. [58]) recognise that pretence understanding is a fundamental part of pretence production.

On the basis of the formalisations proposed, the dissociation hypothesis – as I conceive it applied to pretend play – allows to explain why object substitution is harder for children with autism. Furthermore, it fits in well with both Kavanaugh and Harris [62] and Bigham [8] results. In agreement with Kavanaugh and Harris, it hypothesises that the comprehension of imagination of absent properties or entities should be unimpaired in children with autism relative to controls. The formalisation, however, also suggests that in cases of object substitution, or where the shape of the prop used differs from the one of the imagined object, children with autism will display more difficulties in both pretence understanding and comprehension than controls. Finally, the theory I advanced of pretend play in autism offers an explanation of why elicited pretence is relatively unimpaired in autism (in terms of production of the appropriate answer or action). As it was hinted above, the theory also introduces a novel hypothesis concerning the repetitive and stereo-typical nature of pretence production observed in children within the *ASD*.

As a future extension of the thesis, the analysis of levels of difficulty involved in pretend play could be augmented by augmenting the hypothesis used in this writing, according to which the prototypical features of an object typically guide children's understanding of its function. One could hypothesise that the latter are more used when children detect or guess what is happening within the pretend play episode on the basis of the actions carried out by the experimenter and by typical actions related to a certain object (e.g. bringing an object to the mouth and moving the mandible up and down to chew as related to food items). I would further suggest that neuro-typical children, after the age of 4, seem to be strongly clued by actions and events. On the basis of Bigham [8], it seems that children with autism encounter more difficulties than controls in inhibiting the prototypical features of the object such as shape. This would explain why children with autism encounter more difficulties in pretend plays where the function of the prop is unrelated with its shape. Therefore, an interesting extension of this project might draw a precise and detailed hierarchy of types of pretend play, on the basis of their logical form – i.e. overcoming a contrast between two clauses, augmenting reality with an absent object or property, importance of relational properties, importance of properties expressing prototypical features –, and test whether these levels correspond to different performance on matched tasks.

4.7 Conclusion

Within this chapter I have displayed how the logic for planning elaborated in van Lambalgen and Hamm [115] and van Lambalgen and Stenning [117] can be used to formalise counterfactual imagination, subtractive reasoning and pretend play. In accordance with the theoretical story proposed in chapter 3, I have attempted to exhibit how both subtractive reasoning and pretend play spring from counterfactual imagination.

The proposed model of subtractive reasoning allowed me to identify why children below the age of 12 are not perfect counterfactual reasoners, and rely on basic conditionals to imagine a situation where an event e that happens in reality does not hold true. I sketched a model of proper counterfactual reasoning and subtractive reasoning would look like – on the basis of Abductive

Constraint Logic Programming and the extended version of the Event Calculus introduced in A.3. Moreover, I emphasised which fundamental reasoning processes underlie them, according to the theory I propose.

Finally, I showed how it is possible to formalise pretend play using the logic for planning-Event Calculus-based system [115]. The application of the dissociation hypothesis to pretend play revealed to be extremely fruitful. The generalisation of the dissociation hypothesis to pretend play allowed me to explain many differences in both pretence understanding and production between normally developing and autistic children – differences which correspond to robust experimental evidence. Even more surprisingly, the theory proposed generated a novel explanation of the stereo-typical and repetitive nature of children with autism' pretend play. *Contra* the so-called "generativity hypothesis" (see Jarrols et al. [59]), the present account explains the observed phenomena on the basis of an hypothesis – i.e. the partial impairment in terms of Closed World Reasoning about facts – which already received support from logical inference and neuropsychological studies on autism.

Chapter 5

Conclusions and future work

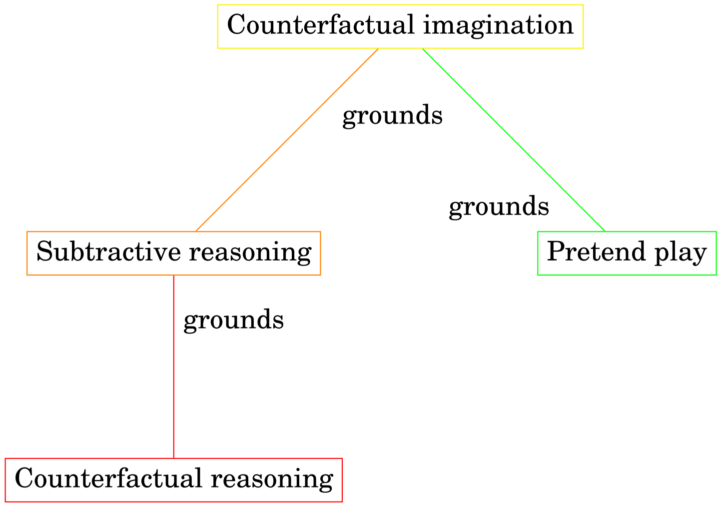
Within the present work, I advanced an analysis of pretend play, focusing in particular on its development in infancy and early childhood. The theory I proposed displayed how the ability to engage in pretend play springs from counterfactual imagination, and how it relates with subtractive and counterfactual reasoning.

On the basis of the analysis of the two theories of pretence proposed by P.L. Harris and by Nichols and Stich, and of the experimental results concerning both counterfactual reasoning and make-believe plays in early childhood, in chapters 3 and 4 I proposed a story on pretend play where theory and logical modelling are highly intertwined. The special union of theoretical and formal analysis presented in this work allowed me to display which reasoning patterns ground both pretend play and subtractive reasoning, and symmetrically which elements distinguish the two.

The story I supported explained pretend play without resorting to the "theory of mind-hypothesis", and it proved that it is possible to account for make-believe plays in terms of offline simulations. The theory I proposed in chapters 3 and 4 combines insights from the Theory Theory and the Simulation Theory families, displaying that the two approaches can be reconciled in a unified framework. Furthermore, I proposed that the *dissociation hypothesis* for subjects with autism, originally formulated on the basis of logical inferences' tasks, generalises to the case of pretend play. The third and fourth chapters probed this novel path for analysing children with autism' performance in pretend play. The hypothesis proved to be extremely fruitful: it fit in well with the majority of the results on pretend play production and understanding. Moreover, it quite naturally led to a novel explanation of the lack of spontaneity and creativity that is often observed in children with autism's pretence.

Finally, the system just introduced can be understood as an extension and improvement of the

Figure 5.1: A developmental hierarchy



theories analysed in chapter 2, attempting to address all the objections posed to Nichols and Stich, and P. Harris. My proposal made explicit in which manner the import of real beliefs and knowledge into a pretend description can be achieved. It also distinguished between pretence understanding and (counterfactual) imagination, leaving open to future empirical investigation the question as to whether imagination is effectively impaired in subjects with autism, or if the apparent difficulties in imagining are due to the previously discussed differences of autistic children's pretend play relative to controls. The story of make-believe plays I advanced, moreover, clarified at which stages in the connection between a pretence representation and the generation of an action the children with autism differ from neuro-typical children. As the reader will remember, at the end of section 2.4.2 the relationship between imagination and counterfactual reasoning appeared extremely unclear. My approach addressed this issue, by proposing a hierarchy of more difficult reasoning processes based on developmental evidence and logical formalisation. As displayed in figure 5.1, the story I argue for in this work envisages counterfactual imagination as the origin from which both subtractive reasoning and pretend play's abilities stem. Full-fledged counterfactual reasoning – or "the sense for the nearest possible world" sought by Perner and colleagues – is a further development of subtractive reasoning.

5.1 Possible extensions and future work

The formalisations here proposed can be augmented in multiple manners. The most obvious extension would be to propose a logical modelling for all different types of pretend play, while in the present work I was able to only sketch how imagination of absent objects and properties would look like according to the theory. Furthermore, as I mentioned in chapter 4.6.1.1, an even more detailed study of pretend play in autism could probe the extend to which children with autism – and children above 3 years – rely on prototypical features in order to understand the function of an object. At the same time, I would be extremely interesting to investigate whether children with autism display a general difficulty in using relational properties, such as actions carried out by a caretaker or experimenter, to understand the function of an object. Within this thesis, I suggested that this phenomenon in pretend play can be explained in terms of the dissociation hypothesis, but further work is required to investigate the extend to which this hypothesis generalises. Notice, *en passant*, that this possible extension of the work might have consequences on the study of analogy in autism, for analogy could be analysed in terms of relational similarities.

In section 4.5, I mentioned that the Cyc knowledge base could be employed together with the Event Calculus, so as to check how such a huge *corpora* would relate to the assumptions made in this work about general knowledge. Moreover, the usage of the Cyc knowledge base – once an efficient translation from the language used in Cyc to Logic Programming is found – would allow to generate a much more general and comprehensive theory of general and context-specific knowledge.

The theory I advanced identifies two necessary requirements for pretence: the comprehension that pretend play episodes have temporal boundaries, and the understanding that pretended actions and events have different effects than real actions. As far as the latter is concerned, the story presented suggests that children should understand the diversity in effects between a pretended transformation and a real one. Therefore, in appendix B I describe a non-verbal task for pretence understanding, which aims at probing whether children discern between real and pretended effects of an event or action. The non-verbal nature – or reduced linguistic complexity – of the task makes for a favourable condition for testing both really young children (as young as 18 months) and children with language impairment, e.g. with autism. The stories I propose in appendix B would furthermore probe children with autism' understanding of symbolic play, so as to corroborate or contrast the results in Bigham [8].

In section 3.2.2, I hinted at a possible extension of the semantics I used. Robert Stärk proposed a four valued semantics for Logic Programming which would permit one to directly model that an agent imagines contradictory information. Stärk proved [108] that the consequences of a program under his four valued semantics are the same – as long as there is no identity– as with Fitting's three valued semantics. I haven't been able to investigate how the four valued semantics proposed by Stärk would behave under the van Lambalgen and Stenning operator (see appendix A). This would however constitute an interesting formal extension of the present work, which may allow to augment the present account with the representation of logical impossibilities.

The present work underlines a fundamental research direction in Logic Programming: the complex logical formalism used in this work and in van Lambalgen and Hamm [115] hasn't yet found a perfect implementation. In particular, the combination of the presence of negative existentials in the body of the Event Calculus and recursion means that, even when Prolog implementations are augmented with integrity constraints, the computational time (or running time) increases exponentially[1]. Therefore, a future extension of the work may consist in finding (or devising) a perfect implementation of the formalism here used.

Finally, as I mentioned in the introduction to the work, I am convinced that the investigation of imagination and pretend play is essentially connected with the metaphysical discussion of the relationship between imagination and modality. If the present work has not offered any metaphysical argument for or against one metaphysical position, this is due to the conviction that a theory proposing a precise and clear understanding of the phenomena under discussion must be as detached as possible from the author's pre-theoretical intuitions. This means, for instance, that the facility with which I have allowed children to imagine impossibilities, facility which I guaranteed in this work, is now open to theoretical and metaphysical discussion. Authors claiming that the negation of a necessary *a posteriori* truth cannot be imagined may argue that the imagination displayed by children does not *really* contradict their claim, because of a mis-identification of what is imagined. Opponents to this principle may rebut that not all types of imagination of impossibilities described seem to be passable of mis-identification. In any case, as a methodological stance, I am convinced that postponing this type of discussions to when the development of the phenomenon under discussion is clearly understood turns out to be very fruitful.

[1]In general, many Prolog implementations use the Martelli-Montanari (see Martelli and Montanari [80]) algorithm for first order unification, which generates the most general unifier. To unify a variable x with a term strictly containing it, e.g. $t(x)$, one would need to execute an additional check (called occurs check). Because the occurs check slows the program, it is often skipped. It is possible, however, to force the program to execute the occurs check, but this addition does not resolve the need for a perfectly matching implementation for the logic here used, for the occur check rule would make the computational time grow excessively.

APPENDIX

A

APPENDIX A: SOME LOGICAL BACKGROUND

Various formal notions need to be introduced in order to obtain a formalisation of pretence that is expressive enough. I have decided to enclose all the details that have been already proved or discussed in the literature in this chapter, with the aim of providing the necessary background to any type of reader.

A.1 Introduction

Logic Programming is a quite fertile formal system, that has been widely studied in Informatics and Artificial Intelligence. It has been used to model abduction, induction, deductive reasoning, knowledge bases, machine ethics, to mention only but a few. Since the variety of approaches within the family of logic programming logics is majestic, I henceforth only introduce the definitions and proofs that are fundamentally needed for the formalisation of pretend play. I mention, where possible, some literature to which the avid reader can be referred.

The introduction of the formal details starts from predicate logic programming. In chapter A.2 the syntactic and semantic details for first order logic programs is displayed. Chapter A.3 introduces some necessary definitions of the event calculus, as it was defined in van Lambalgen and Hamm [115]. I then introduce abduction in logic programming, in chapter A.4.

A.2 Predicate Logic Programming

The material exposed in this section aims at briefly introducing the syntax and semantics for First Order Logic Programming that will constitute the basic framework of this work. The following pages, therefore, cannot represent an exhaustive and complete overview of the many different approaches present in the literature. Finally, the definitions and proofs that will be henceforth introduced have been heavily inspired by the work of Ramli [94], van Lambalgen and Stenning [117] and Lloyd [78].

Let us start with defining an alphabet A consisting of finite or countable infinite disjoint sets of constants, function symbols and predicate symbols, an infinite set of variables, the connectives $\neg, \vee, \wedge, \leftarrow, \leftrightarrow$ and punctuation symbols "(", ",", ")". A also contains the symbols $\top$ denoting a valid formula, and $\bot$ denoting an unsatisfiable formula. We define a language $\mathscr{L}$ on the basis of an alphabet A.

Definition A.1. (Term) Terms are defined inductively as follows:

1. a variable is a term;

2. a constant is a term;

3. if f is a n-ary function symbol and $t_1, \ldots, t_n$ are terms, then $f(t_1, \ldots, t_n)$ is a term.

Following a convention in Logic Programming, I will use upper case letters to denote variables, and lower case letters to denote constants, function and predicate symbols.

A term that does not contain variables is called *ground*.

Definition A.2. (Literal)
A literal is an atom (positive literal) or the negation of an atom (negative literal).

Definition A.3. (Ground term and expressions) Let $Var(t)$ denote the set of all variables occurring in the term t.A variable-free term (i.e. a term t for which $Var(t) = \emptyset$) is called *ground*.

An expression without variables or quantifiers is called *ground*.

Definition A.4. (Well-formed formulas) A well formed formula is defined inductively as follows:

1. If p is an n-ary predicate symbol and $t_1, \ldots, t_n$ are terms, then $p(t_1, \ldots t_n)$ is a formula (called atom);

2. if φ and ψ are formulas, then so are $\neg\varphi$, $\varphi \wedge \psi$, $F \vee G$, $\varphi \rightarrow \psi$.

3. if φ is a formula and x is a variable, then $\forall x \varphi$ and $\exists x \varphi$ are formulas.

A.2.1 Syntax of Logic Programming

Logic Programming is a declarative and relational style of programming, based on first order or propositional logic. The syntax of logic programs is more restrictive than classical first order logic in that the occurrence of implications ($\rightarrow$) is limited to formulas $\varphi_1, \ldots, \varphi_n \rightarrow A$ where A is a literal, and in that iterations of $\rightarrow$s are not allowed. Furthermore, clauses are only allowed to be in the Conjunctive Normal Form, and existential quantifiers are subject to Skolemisation[1].

[1] For the details, see Kowalski [65], chapter A1.

The reasons that lead me to adopt the non-monotonic non-standard formalism to model human reasoning include the nice property that the family of logic programming languages exploit the existence of tractable fragments of predicate logic. The syntactic restrictions of the language, moreover, allow for an efficient proof search, such that derivations are also computations[2]. But there are many more reasons for appreciating the logic for planning introduced in van Lambalgen and Stenning [117]: a model of a consistent set of sentences can be constructed directly from the input. Furthermore, it is in principle possible to perform these computations on neural nets[3]. Within this chapter, I will provide the basic syntactic definitions.

Definition A.5. (Clause) A clause is a finite set of literals of the form

$$\forall x_1, \ldots x_n (L_1 \vee L_n)$$

s.t. $x_1, \ldots, x_n$ are all the variables occurring in $L_1 \vee \cdots \vee L_n$.

Definition A.6. (Definite Clause) A definite program clause (or rule) is a clause of the form

$$A \leftarrow B_1, \ldots, B_n$$

which contains precisely one positive atom A (alternatively, one atom in its consequent). A is called the *head* and $B_1, \ldots, B_n$ is called the *body* of the program clause.

So a rule with head A and negative literals $\neg B_1, \ldots, \neg B_m$ will be written as

$$A \leftarrow B_1, \ldots, B_m$$

or in first order notation this would be

$$B_1 \wedge \cdots \wedge B_m \rightarrow A$$

.

Definition A.7. (fact) A rule (A.5) with an empty body $A \leftarrow$ is called a *fact* or *unit clause*.

Definition A.8. A rule with no positive literal $\leftarrow B_1, \ldots, B_n$ is called a *goal* or a *query*. Goals will be usually denoted by $?B_1, \ldots, B_n$.

Definition A.9. (Horn sentence) A *Horn sentence* is a universally quantified disjunction of literals of which at most one is positive. Thus, goals and rules are Horn.

Definition A.10. (Program) A (definite) program is a finite set of (definite) rules.

Definition A.11. (Definition) In a definite program, the set of all program clauses with the same predicate symbol p in the head is called the *definition* of p.

In what follows, I will assume that the language $\mathscr{L}$ underlying a program $\mathscr{P}$ contains precisely the predicate, function and constants symbols occurring in $\mathscr{P}$.

[2]By the slogan "derivations are also computations" I mean that if there is a derivation of $\varphi(x)$, the derivation will also provide a witness for x. Cfr. van Lambalgen and Stenning [117].

[3]See van Lambalgen and Stenning [117], chapter 8.

A.2.2 Semantics of Logic Programming

The declarative semantics of a logic program is given by a model-theoretic semantics of the formulae in the underlying language. I will hence introduce interpretations and models, focusing in particular on Herbrand interpretations. This approach, roughly speaking, allows us to treat first order atomic formulas as atomic propositions.

Definition A.12. (Herbrand Universe) Let $\mathscr{L}$ be a first order language. The *Herbrand universe* $U_{\mathscr{L}}$ for $\mathscr{L}$ is the set of all ground terms, which can be formed out of the constants and function symbols appearing in $\mathscr{L}$. (In case $\mathscr{L}$ has no constants, we add some constants, say a, to form ground terms.) The Herbrand universe for the language underlying a program $\mathscr{P}$ will be denoted by $U_{\mathscr{P}}$.

Definition A.13. (Ground Instance) A *ground instance* of a formula φ is any ground formula that results from φ by substituting all variables by terms in $U_{\mathscr{L}}$. We denote by $ground(\mathscr{P})$ the set of all ground instances of clauses in program $\mathscr{P}$.

Even though a program $\mathscr{P}$ is finite, as seen in definition A.10, $ground(P)$ may be infinite. To deal with unification issues, $ground(P)$ can be treated as a substitute for $\mathscr{P}$[4]. An alternative approach is to add a constraint solver: more details of this framework will be introduced in chapter A.3. For the moment, for simplicity's sake, I will follow the approach of Ramli [94].

Definition A.14. (Herbrand base) The *Herbrand base* $B_{\mathscr{L}}$ for a language $\mathscr{L}$ is the set of all ground atoms that can be formed by using predicate symbols from $\mathscr{L}$ and ground terms from $U_{\mathscr{L}}$ as arguments. By $U_{\mathscr{P}}$ we denote the Herbrand base for the language underlying the program $\mathscr{P}$.

Let's now introduce three-valued interpretations and models for logic programs. Following van Lambalgen and Stenning [117], we adopt the strong Kleene's definition of truth values for connectives.

Definition A.15. (Interpretation) An *interpretation* I of a program $\mathscr{P}$ is a mapping from the Herbrand base $B_{\mathscr{P}}$ to the set of truth values $\{\top, \bot, u\}$. We represent interpretations by pair $\langle I^{\top}, I^{\bot}\rangle$, where the set $I^{\top}$ contains all atoms which are mapped to $\top$, the set $I^{\bot}$ contains all atoms which are mapped to $\bot$ and $I^{\top} \cap I^{\bot} = \emptyset$. We say an interpretation I is *total* if $I^{\top} \cup I^{\bot} = B_{\mathscr{P}}$.

Definition A.16. (Truth value) The logical value of ground formulae can be derived from the truth tables in fig A.1[5] . A formula φ is true under interpretation I, denoted by $I(\varphi) = \top$, if all its ground instances are true in I. φ is false under I, denoted by $I(\varphi) = \bot$, if there is a ground instance of φ that is false in I. Otherwise is undefined under I, denoted by $I(\varphi) = u$. Two formulae

[4] See Ramli [94].

[5] The biconditional in fig A.1 follows Łukasiewicz's. This definition captures the intuition that $\varphi \leftrightarrow \psi$ is true iff the truth value of φ is the same as the truth value of ψ, and false otherwise.

Figure A.1: Strong Kleene's truth tables (and Łukasiewicz biconditional)

		$\wedge$	$\vee$	$\rightarrow$	$\leftrightarrow$
$\top$	$\top$	$\top$	$\top$	$\top$	$\top$
$\bot$	$\top$	$\bot$	$\top$	$\top$	$\bot$
u	$\top$	u	$\top$	$\top$	$\bot$
$\top$	$\bot$	$\bot$	$\top$	$\bot$	$\bot$
$\bot$	$\bot$	$\bot$	$\bot$	$\top$	$\top$
u	$\bot$	$\bot$	u	u	$\bot$
$\top$	u	u	$\top$	u	$\bot$
$\bot$	u	$\bot$	u	$\top$	$\bot$
u	u	u	u	u	$\top$

	$\neg$
$\top$	$\bot$
$\bot$	$\top$
u	u

φ and ψ are said to be semantically equivalent, denoted by $\varphi \equiv \psi$, if φ and ψ have the same truth value under all interpretations.

Definition A.17. (Model) Let I be an interpretation, I is a model of a formula φ if $I(\varphi) = \top$. For a program $\mathscr{P}$, we say I is a model of $\mathscr{P}$ if I is a model for every clause in $\mathscr{P}$. Similarly, for a set of clauses $\mathscr{C} \subseteq \mathscr{P}$, we say I is a model of $\mathscr{C}$ if I is a model for every clause in $\mathscr{C}$.

Finally, it's possible (and convenient) to define an ordering among interpretations.

Definition A.18. (Ordering among Interpretations) Let $I = \langle I^\top, I^\bot \rangle$ and $J = \langle J^\top, J^\bot \rangle$ be two interpretations. We write $I \subset J$ iff $I^\top \subset J^\top$ and $I^\bot \subset J^\bot$. We write $I \subseteq J$ iff $I^\top \subseteq J^\top$ and $I^\bot \subseteq J^\bot$. If $\mathscr{I}$ is a collection of interpretations, then an interpretation $I \in \mathscr{I}$ is called *minimal* in $\mathscr{I}$ iff there is no interpretation $J \in \mathscr{I}$ s.t. $J \subset I$. An interpretation is called *least* in $\mathscr{I}$ iff $I \subseteq J$ for any interpretation $J \in \mathscr{I}$. A model $\mathscr{M}$ of a program $\mathscr{P}$ is called minimal (resp. least) if its minimal (resp. least) among all models of $\mathscr{P}$.

A.2.3 Program Completion

Definition A.19. (Program Completion) Let $\mathscr{P}$ be a logic program. Consider the following transformation:

1. Replace all clauses in $ground(\mathscr{P})$ with the same head (ground atom) $A \leftarrow Body_1, A \leftarrow Body_2, \ldots$ by the single expression $A \leftarrow Body_1 \vee Body_2 \vee \ldots$, i.e. the *definition* of A (def. A.11).

2. If a ground atom A is not the head of any clause in $ground(P)$ then add $A \leftarrow \bot$.

3. Replace all occurrences of $\leftarrow$ by $\leftrightarrow$.

The resulting set of formulae is called *completion* of $\mathscr{P}$ and is denoted by $comp(\mathscr{P})$. One should observe that in step 1, there may be infinitely many clauses with the same head resulting in a countable disjunction.

A.2.4 The SvL Consequence Operator

Let $\mathscr{P}$ be a logic program. A consequence operator of $\mathscr{P}$ is a function which maps an interpretation of $\mathscr{P}$ to another interpretation of $\mathscr{P}$. It expresses the consequences of $\mathscr{P}$ when the bodies of its clauses are interpreted under the input interpretation.

Definition A.20. (Stenning and van Lambalgen Immediate Consequence Operator) Let I be an interpretation and $\mathscr{P}$ be a general program. The Stenning and van Lambalgen immediate consequence operator is defined as follows:
$\Phi_{\mathrm{SvL},\mathscr{P}}(I) = \langle J^\top, J^\perp \rangle$, where

$$J^\top = \{A \mid \text{there exists } A \leftarrow Body \in ground(\mathscr{P}) \text{ with } I(Body) = \top\}$$
$$J^\perp = \{A \mid \text{there exists } A \leftarrow Body \in ground(\mathscr{P}) \text{ and for all } A \leftarrow Body \in ground(\mathscr{P}) \text{ we find } I(Body) = \perp\}$$

As proved in Ramli [94], the Stenning and van Lambalgen immediate consequence operator is in general monotonic, but not continuous for general programs. It is continuous for ground programs (or propositional programs).

Furthermore, Ramli [94] proves that

Proposition A.1. *Let C be a complete partial order with the least element $\perp$, f be a monotonic mapping on C, x be the least fixed point of f and*
$x_0 = \perp$,
$x_\alpha = f(x_{\alpha-1})$ *for any non-limit ordinal* $\alpha > 0$,
$x_\alpha = lub(x_\beta \mid \beta < \alpha\}$ *for every limit ordinal* α.
Then for some ordinal γ we find $x = x_\gamma$.

By monotonicity of the SvL operation and proposition A.1, we then know that the SvL operator has a least fixed point that is characterised by transfinite induction: iterating $\Phi_{\mathrm{SvL},\mathscr{P}}$ starting from the empty interpretation must eventually yield a fixed point.

Let M be *least fixed point* of $\Phi_{\mathrm{SvL},\mathscr{P}}$ and let

$M_0 = \langle \emptyset, \emptyset \rangle$;

$M_\alpha = \Phi_{\mathrm{SvL},\mathscr{P}}(M_{\alpha-1})$ for any non limit ordinal $\alpha > 0$;

$M_\alpha = \bigcup_{\beta<\alpha} M_\beta$ for every limit ordinal α.

Then for some ordinal γ, it holds that $M = M_\gamma$. The Stenning and van Lambalgen semantics is given by this least fixed point:

Definition A.21. (SvL Model) The least fixed point of $\Phi_{\mathrm{SvL},\mathscr{P}}$ is called the Stenning and van Lambalgen (SvL) model of $\mathscr{P}$.

A.2.5 Negation

The customary treatment of negation as negation as failure in Logic Programming is based on the application of a closed world assumption to negative facts. The *rationale* behind this approach is the following: if φ is not a logical consequence of a program $\mathscr{P}$, then we may assume that $\neg\varphi$. However, on the basis of first order logic undecidability, there is no algorithm which will take an arbitrary φ and decide in finite time whether φ is a logical consequence of $\mathscr{P}$. So the closed world assumption about facts (CWRf) is generally restricted to those $\varphi \in B_P$ whose attempted proofs fail finitely. For a definite program P, the SLD finite failure set of P is the set of all $\varphi \in B_P$ for which there is a finitely failed SLD-tree for $P \cup \{\leftarrow \varphi\}$, i.e. one which is finite and contains no success branches. The procedural counterpart of the closed world assumption (about facts) is called negation as failure. The negation as failure rule (henceforth NF rule) states: if φ is in the SLD finite failure set of P, then infer $\neg\varphi$. Because the finite failure set is a subset of the complement of the success set, negation as failure is generally less powerful than the CWRf. The formal details such as that the declarative and the procedural notions of finite failure set and SLD finite failure set correspond are not essential for our current aims, hence the reader who wishes for more can consult Lloyd [78][6].

When negative information is allowed to appear in the body of a clause, it becomes necessary to extend the definition of derivation, obtaining SLDNF derivations.

Definition A.22. (SLDNF derivation) Let P be a general logic program and G_0 a general goal. An SLDNF-derivation of G_0 is a finite sequence of general goals.

$$G_0 \underset{c_0}{\rightarrow} G_1 \rightarrow \dots \rightarrow G_{n-1} \underset{c_{n-1}}{\rightarrow} G_n$$

such that for all i smaller than the length of the derivation:

1. the selected literal in G_i is positive and G_{i+1} is derived from G_i and C_i by one step of SLD-Resolution;

2. the selected literal in G_i is of the form $\neg A$, the goal $?-A$ has a finitely failed SLD-tree and G_{i+1} is obtained from G_i by removing $\neg A$ (in which case $C_i = FF$).

[6] It is important to note that the completion as defined in def. A.19 and the minimal model as defined in def. A.21 do not coincide when negation as failure is introduced and treated as in [117, chap. 7]. The difference between minimal model and the program completion is due to the usual restriction that the only formulas to which negation as failure apply are the one that fail finitely. The work of Ramli [94] formally proves the claim above, and proposed some alternative ways of dealing with negation. In any case, within the present work I adopt a different approach to negation following van Lambalgen and Hamm [115] and Stuckley [110], as it will be explained in A.3.3.

A.3 Event Calculus

Now that the general framework for first order programs has been laid out, I will extend this basic approach to include the tools to formalise (closed world) reasoning about causality and abduction. In doing this, I follow in the footsteps the procedure in van Lambalgen and Hamm [115], with some minor modifications. Let me start with closed world reasoning about causality, for which a many-sorted first order logic will be needed. The sorts in the original formalisation ([115]) are (1) a set of individual objects O (2) real numbers representing time T, (3) time-dependent properties P, (4) variable quantities Q, (5) event types E[7]. I will start with the exposition of the original axioms, to then introduce a small extension of them

The best way to grasp the intuitions behind the formalisation of closed world reasoning about

[7] Formally, a many-sorted first order language is defined as follows. Let S be a non-empty set, whose elements are called *sorts*. Let S^+ be the set of all finite non-empty words over S. The elements of S^+ are called *sort types* and are written as tuples. For instance, $s_1,\ldots,s_n \in S^+$ is written $(s_1,\ldots,s_n) \in S^+$ Now, it is possible to define a signature Σ over S.

Definition A.23. (Signature) The underlying signature Σ of a many-sorted first order logic consists of

1. a non-empty set S of sorts;
2. a set of function symbols F;
3. a set R of (non-logical) relation symbols.

such that each element in $F \cup R$ corresponds to a sort type, i.e. there is a function $t : R \cup F \to S^+$, and for every $r \in R \cup F$ then $t(r)$ is a sort type. An element $c \in F$ s.t. $t(c) \in S$ is called a constant symbol.

In addition to the signature Σ, the following additional symbols are introduced:

1. the set V of variables, such that each sort $s \in S$ corresponds to a countably infinite subset $V_s \subseteq V$ of variables. In other words, there is a function $v : V \to S$, such that for each $sinS$, $v^{\text{‚àí1}}(s)$ is countably infinite. For each variable $x \in V$, its sort is defined to be $v(x)$;
2. logical predicates: $\wedge, \neg, \forall$;
3. the equality symbol $=$;
4. the left and right parentheses: $($, $)$.

Using Σ and the additional symbols above, we can build terms inductively as follows:

1. each variable $x \in V$ is a term of sort $v(x)$.
2. if $f \in F$ is a function symbol of sort type $(s_1,\ldots,s_n)$ and for each $i = 1,\ldots,n_1$, t_i is a term of sort s_i then $f(t_1,\ldots,t_{n-1})$ is a term of sort s_n.
3. All terms are build in this way.

Finally, from the terms, we inductively define formulas:

1. if t_1 and t_2 are terms of the same sort, then $(t_1 = t_2)$ is a formula.
2. if $r \in R$ is a relation symbol of sort type $(s_1,\ldots,s_n)$ and for each $i = 1,\ldots,n$ t_i is a term of sort s_i then $r(t_1,\ldots,t_n)$ is a formula.
3. if α is a formula, then $\neg\alpha$ is a formula.
4. if α, β are formulas, then $(\alpha \vee \beta)$ is a formula.
5. is α is a formula and $x \in V$ is a variable, then so it $\forall x(\alpha)$.
6. all the formulas are formed in this way.

The signature Σ, additional symbols, and terms and formulas subsequently defined constitute what is known as the many-sorted language $L = L(\Sigma)$ on Σ.

causality is to introduce its axioms providing a informal sketch of their meanings. Therefore, I will now introduce the 5 axioms of the Event Calculus, as formulated in van Lambalgen and Hamm [115], and their extensions for spatial fluents. The *ensemble* of these axioms will be called EC. Let f be a fluent (time-dependent property), t, t', s, r be times, e an event token:

Axiom 1. $Initially(f) \rightarrow Holdsat(f, 0)$

If initially a time-dependent property holds, than this time-dependent property holds at time 0.

Axiom 2. $Holdsat(f, r) \wedge r < t \wedge \neg\exists s < rHoldsat(f, s) \wedge \neg Clipped(r, f, t) \rightarrow Holdsat(f, t)$

A time-dependent property f holds at a certain time t, if f was true at some time r before t and at no earlier than r point in time f was true, and there is no event that causes f to end between r and t.

Axiom 3. $Happens(e, t) \wedge Initiates(e, f, t) \wedge t < t' \wedge \neg Clipped(t, f, t') \rightarrow Holdsat(f, t')$

A time-dependent property f holds at a certain time t', if at time t before t' an event e happens and the event e initiates the time-dependent property f and there is event that causes f to end between t and t'.

Axiom 4. $Happens(e, t) \wedge Initiates(e, f_1, t) \wedge t < t' \wedge t' = t + d \wedge Trajectory(f_1, t, f_2, d) \wedge \neg Clipped(t, f_1, t') \rightarrow Holdsat(f_2, t')$.

A time-dependent property f_2 holds at a certain time t' if an event e happens at time t before t', and the event e initiates an other time dependent property f_1 and f_1 holding between t and d entails that f_2 holds at $t + d$ and $t' = t + d$ and there is no event that causes f_1 to stop between t and t'.

Axiom 5. $Happens(e, s) \wedge t < s < t' \wedge (Terminates(e, f, s) \vee Releases(e, f, s)) \rightarrow Clipped(t, f, t')$

An event causes a time-dependent property f to stop between t and t', if an event e happens at time s, for s between t and t', and the event e either terminates instantaneously f at s or it causes the time-dependent property to change little by little.

The intuitive idea behind axiom 5 of the EC is that $Terminates$ captures the instantaneous change caused by an event, while $Releases$ captures the idea that a variable property may change in absence of any co-occurring event.

$**$ The extension of the EC I will use in chapter 4, augments the five sorts of with a set of tuples of real numbers representing space S. Formally, the axioms of the EC need to be changed, for the sake of clarity I will call this extension EC^*. The axioms $1-5$ of the EC are replaced by the following 5 axioms, where f is a fluent (time-dependent property), t, t', s, r are times, e an event token, and l, l' and g spatial locations:

Axiom 6. $Initially(f,l) \rightarrow Holdsat(f,l,0)$.

If initially a time-dependent property holds true at a certain location l, than this time-dependent property holds true at location l at time 0.

Axiom 7. $Holdsat(f,l,r) \wedge r < t \wedge \neg\exists s < r(Holdsat(f,l,s)) \wedge \neg\exists l' \neq l(Clipped(l,r,f,l',t)) \rightarrow Holdsat(f,l,t)$

If a time-dependent property is instantiated at a certain time r and location l, and at no point it time before r the property f was at location l, and there is no event that ceases the instantiation of f at location l between r and t, then f is still instantiated at l at time t.

Axiom 8. $Happens(e,l,t) \wedge Initiates(e,f,l,t') \wedge t < t' \wedge \neg\exists l' \neq l(Clipped(l,t,f,l',t')) \rightarrow Holdsat(f,l,t')$

A property f is instantiated at l at t' if an event e happens at a certain place g, and e initiates the property f to be instantiated at l at a later time t' and there is no location l' s.t. the property f ceases to be instantiated at l between time t and time t'.

Axiom 9. $Happens(e,l,t) \wedge Initiates(e,f_1,l,t) \wedge t < t' \wedge t' = t+d \wedge l < t' \wedge l' = l+g \wedge Trajectory(f_1,l,t,f_2,g,d) \wedge \neg Clipped(l,t,f_1,l',t') \rightarrow Holdsat(f_2,l',t')$

A time-dependent property f_2 is exemplified at a certain space l' at a certain time t' if an event e happens at location l at time t before t', and the event e initiates an other time dependent property f_1 at l and f_1 holding between t and d entails that f_2 holds at $t' = t+d$ and $l' = l+g$ and there is no event that causes f_1 to stop between t, l and t',l'.

Axiom 10. $Happens(e,g,s) \wedge t < s < t' \wedge l \leq g \leq l' \wedge (Terminates(e,f,g,s) \vee Releases(e,f,g,s)) \rightarrow Clipped(l,t,f,l',t')$

An event e at location g at time s causes a time-dependent property f to cease holding between t and t' and l and l' , for g between l and l', and s between t and t', and the event e either terminates f holding at l,t instantaneously or it causes the property to change little by little.

The extension is based on the following axiom scheme, which will be an instantiation of the truth predicate T_n described in section A.3.7, for $n = 2$. Suppose that a parametrised fluent derives from a formula $\varphi(l,t)$ by nominalisation, resulting in the object $\varphi[\hat{l},\hat{t}]$. Then we can substitute a particular value for l, according to the following axiom scheme:

$$Holdsat(f[\hat{l},\hat{t}],t') \rightarrow Holdsat(f,l',t')$$

If a fluent has as an object a certain location l and it's true at a certain time t, then the time-dependent property is true at that location l at that time t. This roughly means that it is possible to export spatial information from the fluent to an object of $Holdsat$ which now gets defined for arity 3. I will now proceed with the exposition of the general case of the EC as developed in van Lambalgen and Hamm [115], the modifications just introduced are simple modifications of the axioms, hence I will keep the exposition simple by displaying the formalities for the original axioms. The reader doubting of the simplicity of the extensions of the system for this case, will find an argument for its consistency with the EC in van Lambalgen and Hamm [115, p. 82]. $**$

Informally, $Holdsat(f,t)$ captures the idea that fluent f is true at time t, hence it has a similar function as a truth predicate. As it will be seen later on (section A.3.5), one important constraint on the construction of minimal models will be that they capture this intended meaning of $Holdsat$.

Definition A.24. (State) A state $S(t)$ at time t is a first order formula built from

1. Literals of the form $(\neg)Holdsat(f,t)$, for t fixed and possibly different f;
2. equalities between fluents and terms, and between event terms.
3. formulas in the language of the structure $(\mathbb{R},<;+,\cdot,0,1)$.

$**$ For the introduced modification, a state is defined as:

Definition A.25. (State) A state $S(l,t)$ at time t, location l is a first order formula built from

1. Literals of the form $(\neg)Holdsat(f,t,l)$, for t and l fixed and possibly different f;
2. equalities between fluents and terms, and between event terms.
3. formulas in the language of the structure $(\mathbb{R},<;+,\cdot,0,1)$, or $(\mathbb{R}^2,<;+,\cdot,0,1)$.$**$

A.3.1 Scenarios

The EC provides a general framework to reason about causality in general, but it does not suffice to state specific causal relations in a specific situation. For instance EC does not specify how the action of drawing a circle is causally related to a circle being constructed. For doing this, "micro-theories" need to be introduced, with the aim of making explicit the causal relationship holding in a situation.

Definition A.26. A scenario is a conjunction of statements of the form:

1. $Initially(f)$;
2. $S(t) \rightarrow Initiates(e,f,t)$;
3. $S(t) \rightarrow Terminates(e,f,t)$;
4. $S(t) \rightarrow Happens(e,t)$;

5. $S(t) \rightarrow Releases(e,f,t)$;

6. $S(f_1,f_2,t,d) \rightarrow Trajectory(f_1,t,f_2,d)$.

where $S(t)$ or, more generally $S(f_1,f_2,t,d)$, is a state in the sense of definition A.24

Definition A.27. (Definition of event) A definition of an event is a statement of the form $\varphi \rightarrow Happens(e,t)$, where φ contains only $Happens$ formulas, and e does not occur in φ. A definition of a fluent is defined similarly, with $Holdsat$ substituted for $Happens$.

$**$

Definition A.28. A scenario is a conjunction of statements of the form:

1. $Initially(f)$;

2. $S(l,t) \rightarrow Initiates(e,f,l,t)$;

3. $S(t) \rightarrow Terminates(e,f,l,t)$;

4. $S(t) \rightarrow Happens(e,l,t)$;

5. $S(t) \rightarrow Releases(e,f,l,t)$;

6. $S(f_1,f_2,l,t,l',d) \rightarrow Trajectory(f_1,l,t,f_2,l',d)$.

where $S(l,t)$ or, more generally $S(f_1,f_2,l,t,l',d)$, is a state in the sense of definition A.25

Definition A.29. (Definition of event) A definition of an event is a statement of the form $\varphi \rightarrow Happens(e,l,t)$, where φ contains only $Happens$ formulas, and e does not occur in φ. A definition of a fluent is defined similarly, with $Holdsat$ substituted for $Happens.**$

A.3.2 Constraint Structure

I am now going to extend first order logic programming, as it has been introduced previously, to constraint logic programming, following [115, chap. 5]. The main reason to adopt this formal apparatus is to obtain a symmetrical treatment of positive and negative information about terms substitution. In the standard approach to first order logic programming, indeed, the unification algorithm applied to two atoms determines how two terms are made identical by substituting certain terms for the variables occurring in the atoms. The substitution is executed immediately. In constraint logic programming instead when the unification algorithm determines that a term t should be substituted for a given variable x, the constraint $t = x$ is added without executing any substitution. Let $\{a = b\}$, for two atoms a and b, denote the set of equations between terms which unify a and b if they are unifiable, otherwise the set $\{a = b\}$ is set to $\perp$. This also entails that while in standard logic programming a resolution branch is successful if is terminates with a

$\perp$, in constraint logic programming a successful branch will terminate with some constraints c that are satisfiable wrt the constraint structure. In order to understand what this means, let me define first what a constraint structure is.

Let $\mathscr{K}$ be the language containing the primitive predicates of the Event Calculus (i.e. $Holdsat(f,l,t)$, $Initiates(e,f,t)$, $Terminates(e,f,t)$, $Happens(e,f)$, $Releases(e,f,t)$, $Trajectory(f_1,t,f_1,d)$, $Clipped(t,f,t')$). Let $\mathscr{L} = \{0,1,+,\cdot,<\}$.

Definition A.30. (Constraint Structure) A *constraint structure* $\mathscr{A}$ is constituted by $\mathscr{L}$ extended with the set of all constants, functions, parametrised fluents and events. $\mathscr{A}$ is described by the union of the theories CET and the theory of $\mathscr{L}$.

The language $\mathscr{L}$ is completely axiomatised, but we details of its axiomatisation are not essential for our readers. They can be found in Hodges [50]. The set of these axioms will be called $\mathscr{T}$.

So we have that the primitive predicates of the Event Calculus are axiomatised by EC, the constants, functions denoting objects, and parametrised fluents and events are described by the theory of Clark's equality (i.e. CET), which consists of the following set of statements for function symbols f,g and terms t of the language:

1. $f(Y_1,\ldots,Y_n) = f(Z_1,\ldots,Z_n) \rightarrow Y_1 = Z_1 \wedge \cdots \wedge Y_n = Z_n$;
2. $f(Y_1,\ldots,Y_z) \neq g(Z_1,\ldots,Z_n)$, where f and g are different;
3. if Y occurs in t, $y \neq t$.

Our reader may wonder how the constraint structure just defined will relate to the Event Calculus previously introduced. The programmed predicates of the Event Calculus will have as arguments terms that must be interpreted in the given constraint structure. Henceforth, formulas in $\mathscr{L}$ will be called constraints (denoted by c,c',...), for instance $s < t$ or $t = d + t'$. Furthermore, we can now define the resolution proof system for definite constraint logic programming.

Definition A.31. (Resolution for CLP) Let $?c, b_1,\ldots,b_i,\ldots,b_n$ be a goal and $d_1,\ldots,d_k,c' \rightarrow a$ be a program clause, where c and c' are constraints, then a new goal $?c'', b_1,\ldots,d_1,\ldots,d_k,\ldots,b_n$ can be derived from these two clauses if c'', defined as $c'' = \{c \wedge \{b_i = a\} \wedge c'\}$, is satisfiable in $\mathscr{A}$.

Definition A.32. (Successful, finitely failed and infinite branch) A branch is a derivation tree is *successful* it is finite and ends in a query of the form $?c$, where c is a satisfiable constraint. Note that the query is not allowed to contain an atom. A branch in a derivation tree is *finitely failed* if it ends in a query $?c, b_1,\ldots,b_n$ such that either c is not satisfiable or no program clause is applicable to the b_i. Otherwise the branch is called *infinite.*

The definitions just introduced is still not sufficient to deal with the Event Calculus, because of the occurrence of negations in the bodies of the axioms. Therefore, it is necessary to extend definite constraint logic programming by allowing formulas containing only $\neg$, $\exists$ and $\wedge$; the formulas in our language will hence become classically equivalent to first order formulas of arbitrary complexity.

Definition A.33. (Complex sub-goal) A *complex* sub-goal is defined recursively to be

1. an atom in $\mathscr{K}$;

2. $\neg\exists\overline{x}(c, b_1 \wedge \cdots \wedge b_m)$ where c is a constraint and each b_i is a complex sub-goal[8].

Definition A.34. (Complex Body) A *complex body* is a conjunction of complex sub-goals. A *normal program* is a formula $\varphi \rightarrow a$ of $CLP(\mathscr{T})$ s.t. φ is a complex body and a is an atom.

A.3.3 Negation

The customary treatment of negation as negation as failure has been introduced in A.2.5. Van Lambalgen and Hamm in [115] prefer *constructive negation* to the standard negation as failure approach, on the basis of the following motivation: with NF negative queries only yield to the answers "true" o "false", while with constructive negation also negative queries are able to terminate with computed answer substitutions. Within the constraint logic programming framework just introduced, this means that both positive and negative queries can start successful computations ending in constraints[9]. Disregarding the possibility of infinite derivations, van Lambalgen and Hamm [115] propose a simplified version of constructive negation, that is modelled on NF. Let me start by defining a *frontier*

Definition A.35. (Frontier) A $(P,\mathscr{A},R)$ *frontier* of a derivation tree of G is a finite set of nodes in the derivation tree, excluding the root, s.t. every derivation of G is either finitely failed or passes through exactly one frontier-node.

The operational meaning of constructive negation may be given by the extension of CLP resolution for complex sub-goals:

Definition A.36. (Derivation with complex goals) In a $(P,\mathscr{A},R)$-derivation tree for a goal G of the form

$$c, D_1, \ldots, D_j, \ldots, D_n,$$

where D_j is the selected complex sub-goal by the computation rule R, the children of G are given by

1. if D_j is an atom, the children are given by the standard definition;

[8] I use $\overline{(\varphi)}$ to denote finite tuples of objects, e.g. $\overline{x}$ denotes $(x_1, \ldots, x_n)$.

[9] For a fully formal introduction to constructive negation, see Stuckley [110].

2. if D_j is $\neg\exists c_q, \overline{Q}$, where c_q is a constraint and $\overline{Q}$ is a conjunction of complex sub-goals, then let c' be a constraint s.t. $\mathscr{A} \vDash c \to c'$. Let the goal of the negative sub-derivation be $c' \wedge c_1, (B_1 \wedge \cdots \wedge B_m)$. Let $F = \{(c' \wedge c_1, B_1), \ldots, (c' \wedge c_n, B_n)\}$ be a frontier chosen by computation rule R.

   a) If F is empty, then the unique child is the goal

   $$c, D_1, \ldots, D_{j-1}, D_{j+1}, \ldots D_m;$$

   b) otherwise, let c'_i, N_i be constraints and conjunctions of complex sub-goals respectively s.t.

   $$(c'_1 \wedge N_1) \vee \cdots \vee (c'_p \wedge N_p) \leftrightarrow (\neg\exists\overline{x}, \overline{y}_1 c_1 \wedge B_1) \wedge \cdots \wedge (\neg\exists\overline{z}, \overline{y}_m c_m \wedge B_m)$$

   where the $y_k, 1 \leq i \leq m$ denote the variables in $c_k \wedge B_k$ which do not appear in $c \wedge c_q \wedge \overline{Q}$ and each of the $c'_i \wedge N_i$ is a complex sub-goal.There are children

   $$(c \wedge c'_i, D_1, \ldots, D_j, N_i, D_{j+1}, \ldots, D_n)$$

   for each $1 \leq i \leq p$ s.t.$c \wedge c'_i$ is satisfiable. If $p = 0$, there are no children.

A.3.4 Soundness

Definition A.37. A query $?c, G$ is totally successful if its derivation tree includes successful branches ending in constrains $c \wedge c_1, \ldots, c \wedge c_n$ s.t. $\mathscr{A} \vDash \forall\overline{x}(c \to c_1 \vee \cdots \vee c_n)$.

Intuitively, a totally successful query is one that is successful in all its branches, and this notion is much stronger than satisfiability. As it has been done for first order logic programming, a fundamental technical tool is the one of completion.

Definition A.38. (Completion for CLP) Let $\mathscr{P}$ be a normal program, consisting of clauses

$$\overline{B}^1 \wedge c_1 \to p^1(\overline{t}^1), \ldots, \overline{B}^n \wedge c_n \to p^n(\overline{t}^n)$$

where the p^i are atoms. The completion of $\mathscr{P}$, denoted by $comp(\mathscr{P})$, is computed by the following recipe:

(1) choose a predicate p that occurs in the head of a clause of $\mathscr{P}$

(2) choose a sequence of new variables $\overline{x}$ of length the arity of p

(3) replace in the i-th clause of $\mathscr{P}$ all occurrences of a term in $\overline{t}_i$ by a corresponding variable in $\overline{x}$ and add the conjunct $\overline{x} = \overline{t}_i$ to the body; we thus obtained $\overline{B}^i \wedge c_i \wedge \overline{x} = \overline{t}_i \to p^i(\overline{x})$

(4) for each i, let $\overline{z}$ be the set of free variables in $\overline{B^i} \wedge c_i \wedge \overline{x} = \overline{t}_i$ not in $\overline{x}$

(5) given p, let $n_1, \ldots, n_k$ enumerate the clauses in which p occurs as head

(6) define $Def(p)$ to be the formula $\forall\overline{x}(p(\overline{x}) \leftrightarrow \exists\overline{z}_{n1}(\overline{B}^{n1} \wedge c_{n1} \wedge \overline{x} = \overline{t}_{n1}) \vee \cdots \vee \exists\overline{z}_{nk}(\overline{B}^{nk} \wedge c_{nk} \wedge \overline{x} = \overline{t}_{nk}))$

(7) $comp(\mathscr{P})$ is then obtained as the formula $\bigwedge_p Def(p)$, where the conjunction ranges over predicated p occurring in the head of a clause of $\mathscr{P}$.

Theorem 1. *(Soundness) Let $\mathscr{P}$ be a normal program on the constraint structure $\mathscr{A}$ and let $\mathscr{I}$ be the axiomatization of $\mathscr{A}$.*

1. *If the query $?c, G$ is totally successful, then $\mathscr{T} + \mathscr{P} \models \forall\overline{x}(x \to G)$;*

2. *If the query $?c, G$ id finitely failed, then $\mathscr{T} + \mathscr{P} \models \neg\exists\overline{x}(c \wedge G)$.*

A.3.5 Minimal Models

As it was the case for first order logic programming (chapter A.2), intuitively a model for a program should make true only what is forced by the formulas in the program. The fact that minimal models make true only what is forced by the program nicely mirror the closed world assumption – which is the driving *rationale* behind the present approach – according to which all the information and events that are not directly entailed by the program is assumed to be false. Within the formalism proposed here, in van Lambalgen and Stenning [117] and van Lambalgen and Hamm [115], minimal models do not grow linearly, in the sense that even if it is assumed that everything not forced to occur by the scenario is assumed to be false, later additions may change this. For instance, the processing of a discourse leads to a non-monotonic progression[10].

In the framework of CLP and of the Event Calculus, this intuitive idea shall be extended to: a minimal model should make true only what is required by the scenario and the axioms of the event calculus. Informally, this should mirror the fact that the only events and causal influences that obtain are the one that we observe happening or that are forced to happen by our general rules of causality – i.e. the axioms of the EC–. Extending the definition of minimal model from PLP (see def. A.21) is not trivial at all, because it becomes again necessary to ensure that the minimal models are indeed minimal. Furthermore, it is required from us to check that a scenario (def. A.26) has a unique minimal model.

Let me fix as the domain of the models to be considered in such a way that includes the real numbers together with a finite number of objects, event types and fluents. The general case for Logic Programming tells us that using Klenee's three valued semantics, it is possible to obtain a set of models on the given domain of the completion of the program under consideration that has a *least* model. In order to define the minimal models required for the current logic, some definitions from van Lambalgen and Hamm [115] need to be introduced.

Definition A.39. A subset of $\mathbb{R}$ is *semialgebric* is it is a finite union of sets of the form $\{x \in \mathbb{R}^n \mid f_1 = \cdots = f_k = 0, g_1 > 0, \ldots, g_l > 0\}$ where f_i and g_j are polynomials.

[10] See van Lambalgen and Hamm [115].

Intuitively, fluents should generally determine semialgebric sets and the events should mark the beginning and the end of fluents. However, while in the structure $\{0,1,+,\cdot,<\}$ it is only possible to define semialgebric sets, in normal programs such as $\mathscr{P}$ together with the constraint theory $\mathscr{T}$ it could be that more complex sets are definable. So, following van Lambalgen and Hamm [115] and Stuckley [110], I will introduce some relevant results. However, curious readers are vividly suggested to consult the original works.

Definition A.40. A query $?c, G$ is *finitely evaluable* wrt a program $\mathscr{P}$ if its derivation tree is finite – i.e. all branches in the derivation tree starting from $?c, G$ end either in success or in finite failure. A normal program $\mathscr{P}$ is *finitely evaluable* if every query is finitely evaluable wrt $\mathscr{P}$.

Because a query $?c, G$ may contain variables ranging both over real numbers, and over objects, events and fluents, we will use $\hat{x}$ for the variables over real numbers and the variable $\hat{y}$ for objects, events and fluents. Let $\hat{y}$ be fixed.

Theorem 2. *Let $\mathscr{T}$ be a constraint theory describing the constraint structure $\mathscr{A}$. Let $\mathscr{P}$ be a normal program consisting of the axioms of the event calculus together with a scenario. Let $?G$ be a finitely evaluable query in the language of the event calculus. Let $\hat{b}$ be an assignment to the variable $\hat{y}$. Then, there exists a semialgebric set defined by a constraint $c(\hat{x})$ s.t. $\mathscr{T} + comp(\mathscr{P}) \vDash \forall(G(\hat{x}\hat{b} \rightarrow c(\hat{x}))$.*

Proof. [11] By hypothesis the derivation tree whose top note is $?G$ is finite, hence those terminal nodes which are not marked as failures are marked by a constraint from the language of the structure, i.e. $\{0,1,+,\cdot,<\}$ together with a set of objects, (parametrised) events and (parametrised) fluents, it then follows from lemma 7.3 in Stuckley [110] that there is a constraint $c'(\hat{xy})$ s.t.

$$\mathscr{T} + comp(\mathscr{P}) \vDash \forall(G(\hat{xy} \leftrightarrow c'(\hat{xy})$$

Define $c(\hat{x}) = c'(\hat{x}\hat{b})$, then

$$\mathscr{T} + comp(\mathscr{P}) \vDash \forall(G(\hat{x}\hat{b} \leftrightarrow c(\hat{x}),$$

and an easy extension of Tarski's theorem on "quantifier elimination for real-closed field"[12] shows that $G(\hat{x}, \hat{b})$ represents a semialgebric set. ■

Corollary A.1. *Let $\mathscr{P}$ be as above, and suppose it is finitely evaluable. Then the theory $\mathscr{T}$ + $comp(\mathscr{P})$ has a unique model on $\mathbb{R}$. In this model all the real parts of the primitive predicates are represented by semialgebric sets. Actually, it suffices to require that for all fluents $f(\hat{x})$ in the scenario, the query $?Holdsat(f(\hat{x}), t)$ is finitely evaluable.*

Proof. [13] We prove a stronger statement. Since the scenario is finite, it mentions only finitely many fluents (possibly containing parameters for reals or individuals). By the definition of

[11] From van Lambalgen and Hamm [115], p. 64-65.
[12] See Hodges [50], chapter 8.
[13] As in van Lambalgen and Hamm [115], p. 65.

scenario (def. A.26), in $\mathscr{T}+comp(\mathscr{P})$ every primitive predicate can be defined in terms of $Holdsat$ and relations and functions from the constraint language $\mathscr{L}$.[14] We therefore only have to consider computations involving $Holdsat(f(\hat{x}),t)$. Since the derivation tree for $Holdsat(f(\hat{x}),t)$ is finite, by theorem 2, $\mathscr{T}+comp(\mathscr{P})$ implies that $Holdsat(f(\hat{x}),t)$ is equivalent to a constraint and hence this formula is definable in terms of semialgebric sets and finitely many constants. It then follows from the results of [110, section 6] that the model determined by the answer to the queries $?Holdsat(f(\hat{x}),t)$ for each f is a model of $\mathscr{T}+comp(\mathscr{P}$, which is unique. ■

Corollary A.2. *Let $\mathscr{P}$ be as above, the following are equivalent:*

(a) any model of $\mathscr{T}+comp(\mathscr{P})$ is completely determined by its restrictions to $\mathbb{R}$;

(b) all sets defined by a formula of the form $Holdsat(f(\hat{x},t)$ are semialgebric.

Proof. [15] The direction from (a) to (b) follows from corollary A.1. The converse direction follows from Beth's definability theorem and Tarski's theorem. ■

Finally, there is also a completeness results, of which I only state the theorem. The proof can be found in van Lambalgen and Hamm [115], p. 66.

Theorem 3. *Let $\mathscr{P}$ be as above. The query $?Holdsat(f(\hat{x},t)$ is finitely evaluable if $\mathscr{T}+comp(\mathscr{P})\vDash \forall t(Holdsat(f,t)\leftrightarrow c(t))$ for a constraint c.*

The assumption of finite evaluability is fairly strong, because it entails that the scenarios under consideration are sufficiently complete. However, this assumption

A.3.6 Gödel coding

In chapter 3.5.4 and 4.5.1, I explain how the mental representations concerning the pretend play can be related to the actual knowledge and beliefs of an agent. Furthermore, I hypothesise that only from such a comparison pretend behaviour can originate. In order to achieve such an import from the pretence environment, however, it would be necessary to have first order formulas as the object of a predicate. Obviously, this violates the definition of a well ordered formula in our language. Therefore, in this section I will introduce the device of Gödel coding and the Feferman's method to have binary paring functions. This approach is employed in van Lambalgen and Hamm [115] to formalise nominalisation. With this term I refer to the phenomenon that allow us to use verbal phrases to represent events and time-dependent properties. For example a verb as *run* could be represented in our language by $run(x,t)$. Now, if we have the formula $\exists t(run(x,t))$, this formula formally represents an object. On the contrary, the formula $run[j,\hat{t}]$ will represent an function from time to truth value, for j denoting an individual. Formally, $\exists t(run(x,t))$ is called

[14] Hence it is essential that no primitive predicate occurs both in the head and in the body of the same clause.
[15] See van Lambalgen and Hamm [115], p. 65.

Table A.1: Gödel numbers of symbols

Symbol	(	)	,	$\neg$	$\vee$	$\exists$	$=$	v_i	A_i^n	f_i^n
code	1	3	5	7	9	11	13	$2 \cdot 5^i$	$2^2 \cdot 3^n \cdot 5^i$	$2^3 \cdot 3^n \cdot 5^i$

an event type or a perfect nominal, while $run[j, \hat{t}]$ is a fluent or an imperfect nominal. Roughly speaking, the distinction between perfect and imperfect nominal can be understood as follows: perfect nominalisation maintains temporal information, while imperfect nominalisation loses track of it. The reader shouldn't be too scared: the use I am going to make of Gödel coding is quite limited, and I will henceforth only introduce the formal details that are strictly needed to understand the treatment.

Let $\mathscr{L}_{/}$ be a first order language, extending the language of the reals, and let S_0 be a theory formulated in $\mathscr{L}_{/}$ s.t. S_0 has at least axioms for $+, \cdot$. The language $\mathscr{L}_0$ and the theory S_0 suffices to code formulas as natural numbers: this general strategy is basic idea behind Gödel numbering. Even though the details of the coding actually employed do not matter much for my aims, for the sake of completeness I mention one of the possible strategies for assigning a number to every formula in the language $\mathscr{L}_0$.

Firstly, let me pick one way to give to each symbol a unique number. On the basis of the numbering for symbols in fig. A.1, we can assign code numbers for symbols in the language of arithmetic. Then, the code numbers is extended to all finite sequences of symbols by assigning to an expression E consisting of a sequence of symbols $S_1, S_2, \ldots, S_n$ the code number $\#(E)$ for the sequence $(|S_1|, |S_2|, \ldots, |S_n|)$ according to the scheme for coding finite sequences of numbers by single numbers based on prime decomposition[16]. Let me go through one example taken from Boolos et al. [10] , to exemplify the general strategy:

Example A.1. *The code number for* $0 = 0$*, or better* $= (0,0)$ *is that for the sequence* $|=|, |(|, |0|, \ |,| \ , |0|, |)|$*. From fig. A.1, the code number for* $=$ *is* 13*, for* (*is* 1*, for* 0 *or* f_0^0 *is* $2^3 \cdot 3^0 \cdot 5^0 = 8$*, and so on. Hence, the code number for* $= (0,0)$ *is that for* $(13, 1, 36, 5, 36, 3)$*, which is* $2^6 \cdots 3^{13} \cdots 5 \cdot 7^{36} \cdot 11^5 \cdot 13^{36} \cdot 17^3$.

Now that a rough idea of the basic mechanism behind the tool of coding has been provided, following Feferman [27] it is possible to define a binary pairing function π in $\mathscr{L}_0$, and two unary projection functions π_1 and π_2. Intuitively, these are needed to abstract certain variables $\hat{x}$ in the Gödel code for a formula φ, and uses other variables $\hat{y}$ as parameters.

For the Feferman's extension of Gödel coding, it is necessary to assume that the language under treatment $\mathscr{L}_0$ has a constant symbol 0, a binary operation symbol π and two binary operation symbols π_1, π_2. Moreover, let's assume that the following two statements are provable in the theory S_0 :

[16] This approach also allows to distinguish for a single symbol S between the code number $\#(E)$ of S as an expression from the code number $|S|$ of S *qua* symbol.

1. $\pi(x,y) \neq 0$;

2. $\pi_1(x,y) = x \wedge \pi_2(x,y) = y$.

So π acts as a pairing function from $M \times M$ to $M - \{0\}$, for which π_1 and π_2 work as projection operations. Furthermore, tuples of the form $(\tau_1,\dots,\tau_k)$ can be defined recursively by $\tau_1 = \tau$ and $(\tau_1,\dots,t_{k+1}) = ((\tau_1,\dots,\tau_k),\tau_{k+1})$. One could also introduce the corresponding projection operations $\pi_i^k (1 \leq i \leq k)$ s.t. in S_0 the following formula $\pi_i^k(x_1,\dots,x_k) = x_i$ is satisfied.

Finally, always following the procedure in Boolos et al. [10] and [115, ch.4], it is possible to see the connection between Feferman's operators and Gödel coding. Let φ be a formula of $\mathscr{L}$, and let $\ulcorner\varphi\urcorner$ be the Gödel number of φ or its corresponding numeral in $\mathscr{L}_0$[17]. Furthermore, let φ be a formula with free variables among $x_1,\dots,x_k,y_1,\dots,y_n$. The term $(\ulcorner\varphi\urcorner,y_1,\dots,y_n)$ in $\mathscr{L}_0$ serves as an operation that treats $x_1,\dots,x_k$ as bound variables, while treating the $y_1,\dots,y_n$ as free variables. Because this operation "bounds" the variables $x_1,\dots,x_k$, then it is possible to write the formula according to the following definition.

Definition A.41. $\Delta_k \quad \varphi[\hat{x}_1,\dots,\hat{x}_k,y_1,\dots,y_n] = (\ulcorner\varphi\urcorner,y_1,\dots,y_n)$. For $k = 1$ in Δ_1, we use the set theoretic notation $\Delta_1\{x \mid \varphi(x,y_1,\dots,y_n) = \varphi[\hat{x},y_1,\dots,y_n]$. If $k = 0$, we write $\varphi[y_1,\dots,y_n] = (\ulcorner\varphi\urcorner,y_1,\dots,y_n)$. Finally, if both k and n are set to 0, then we simply write $\ulcorner\varphi\urcorner$.

A.3.7 Finishing touches

As a finishing touch, it would be desirable to have a general method to use the results of nominalisation in $\mathscr{L}_0$ within computations of the EC. For instance, given a fluent $run[j,\hat{t}]$, if it has been derived a set s such that $Holdsat(run[j,\hat{t}],s)$, it would be desirable to be able to conclude that for the set s $run(j,s)$. One could try to do so by the following schema:

$$Holdsat(\varphi[\hat{t}],s) \leftrightarrow \varphi(s)$$

This strategy however leads to contradictions[18]. Roughly speaking, this is due to the fact that the predicate $Holdsat$ has been used for a truth predicate, and that it has been allowed to iterate $Holdsat$ formula within each other.

Proceeding with other, let me start by introducing a truth predicate T_n in $\mathscr{L}_0$, whose intuitive meaning is for $T_n(x_1,\dots,x_{n-1},z)$ the tuple $(x_1,\dots,x_n)$ satisfies (the formula coded by) z. This intuitive meaning is captured by the following axiom:

Axiom 11. *For any formula φ in the language $\mathscr{L} = \mathscr{L}_0 \cup \{T_n \mid n \in \mathbb{N}\}$ and for all n*

$$T_n(x_1,\dots,x_n\varphi[\hat{u}_1,\dots,\hat{u}_n,y_1,\dots,x_n] \leftrightarrow \varphi(x_1,\dots,x_n,y_1,\dots,y_n)$$

[17] Following [115, chap. 4], $\ulcorner\varphi\urcorner$ will be used interchangeably for the term in $\mathscr{L}$ and for the object denoted by that term in a model $\mathscr{M}$ for $\mathscr{L}$.

[18] See van Lambalgen and Hamm [115],chapter 6.

For the case $n = 1$, we have

$$(4) T_1(x, \varphi[\hat{u}, y_1, \dots y_m] \leftrightarrow \varphi(x, y_1, \dots, y_m)$$

Because T_1 is the general case of $Holdsat$, the following results are formulated in terms of it, they can however easily be generalised.

Axiom 12. [19] (4) *can be reformulated as a logic program. The logic program $\mathcal{T}_1$ is characterised by the following clauses:*

1. $T_1(x, \varphi[\hat{u}, \overline{y}]) \wedge T_1(x, \psi[\hat{u}, \overline{y}]) \rightarrow T_1(x, (\varphi \wedge \psi)[\hat{u}, \overline{y}])$;
2. $T_1(x, \psi[\hat{u}, \overline{y}]) \rightarrow T_1(x, \varphi \vee \psi[\hat{u}, \overline{y}])$;
3. $T_1(x, \varphi[\hat{u}, \overline{y}]) \rightarrow T_1(x, \varphi \vee \psi[\hat{u}, \overline{y}])$;
4. $\neg T_1(x, \varphi[\hat{u}, \overline{y}]) \rightarrow T_1(x, \neg\varphi[\hat{u}, \overline{y}])$;
5. $T_1(x, \varphi[\hat{u}, \overline{y}z]) \rightarrow T_1(x, \exists z \varphi[\hat{u}, \overline{y}z])$;
6. $\neg\exists z \neg T_1(x, \varphi[\hat{u}, \overline{y}z]) \rightarrow T_1(x, \forall z \varphi[\hat{u}, \overline{y}z])$;
7. $T_1(x, \varphi[\hat{u}, \overline{y}]) \rightarrow T_1(x, T_1[\hat{v}, \varphi[\hat{u}, \overline{y}])$;

The logic program $\mathcal{T}_1^-$ is $\mathcal{T}_1$ with the conditions (7) *deleted. The logic program $\mathcal{T}_1^-$ is $\mathcal{T}_1$ with the addition of the following clause, for all atomic formulas $A(x, \overline{y})$:*

$$T_1(x, A[\hat{u}, \overline{y}]) \rightarrow A(x, \overline{y}).$$

Firstly, $\mathcal{T}_1$ is a correct constraint logic program in the sense of definition A.34. Furthermore, the fluents occurring on the right hand side of the clauses in 12 are of greater complexity than those occurring on the left side. Hence, van Lambalgen and Hamm [115, p.80] prove the following lemma.

Lemma A.1. *Let $SCEN$ be a scenario in which all fluents are derived from atomic formulas of $\mathcal{L}_0$. Then the completion of the logical program consisting of $SCEN$, the axioms of the Event Calculus and $\mathcal{T}_1^-$ is consistent.*

I omit the proof of the lemma, but its result can be easily seen considering that $\mathcal{T}_1^-$ does not allow iterations of the truth predicate. For the general case, we ensure that computations performed on fluents can be interpreted as applying to real formulas, not only to the fluents derived from them.

[19] See van Lambalgen and Hamm [115, p.79]

Corollary A.3. *A model $\mathscr{M}$ of $SCEN + EC$ which also satisfies ${\mathscr{T}_1}^-$ may be expanded to a model $\mathscr{N}$ of $\mathscr{L}_0$ in the following manner: if a predicate $A(x,t)$ corresponds to a fluent $f(x)$ in the scenario, interpret $A(x,t)$ on $\mathscr{N}$ by the set $\{(x,t) \mid Holdsat(f(x),t)\}$; otherwise let the interpretation of A be arbitrary.*

In this way the validity of axiom 12 is ensured, and hence also the axiom scheme (4). The lemma hypothesises that the scenario only has information about fluents derived from atomic formulae, therefore the behaviour of complex formulae on $\mathscr{N}$ is trivially consistent with the predictions of the scenario. A correspondent consideration can be formulated for the case in which one does not assume that the only fluents occurring in the scenario are those derived from atomic formulas, and it can be found in van Lambalgen and Hamm [115, pp. 80–82].

A.4 Abductive Logic Programming

In the last 50 years, abductive reasoning has become the object of growing interest in Artificial Intelligence. When we face some unexplained observations, an effective strategy to explain them is to generate plausible hypotheses of the observations. This explanation can take the form of integrating general rules or specific facts (see Kowalski [65]). The latter can be understood as introducing plausible causes for the relevant observation.

Let's consider one example that might clarify the matter: *You wake up and see that the grass in your backyard is wet.* On the basis of your knowledge about wet grass, you know that the grass can be wet because it rained or because the sprinkler was on. Both sentences "It rained last night" and "the sprinkler was on" constitute a good explanation of your observation that the grass is wet.

The search of explanations, however, cannot generally be achieved only on the basis of forward or deductive reasoning.

(1) The grass is wet if it rained.
(2) The grass is wet if the sprinkler was on.
(3) The grass is wet.

Classically, affirmation of the antecedent is a fallacy. Also using the non-monotonic logic for planning introduced in van Lambalgen and Stenning [117] and the propositional version as presented in A.2, it is impossible to deduce that either the sprinkler was on or it rained. For the sake of simplicity we show what we would obtain with the propositional version of the logic for planning and only forward reasoning. Let $Holdsat(grass\_wet,t)$ represent "the grass is wet", $\exists S(Holdsat(rain,S), S < T)$ "it rained" and $\exists S(Holdsat(sprinkler\_on,S), S < T)$ "the sprinkler was on". Let's assume that (1) and (2) are the only two available rules about the grass being wet. By completion (see def. A.38), we obtain

$$comp(P) = \Big\{\forall X,Y Holdsat(X,Y) \leftrightarrow (X = wet\_grass \wedge Y = t) \bigvee (X = wet\_grass, Y = t)$$
$$\wedge \exists S(Holdsat(rain,S), S < T) \bigvee (X = wet\_grass, Y = t) \wedge \exists S(Holdsat(sprinkler\_on,S), S < T)\Big\}.$$

Then, it is trivial to check that $comp(P) \vDash_3 Holdsat(grass\_wet, T)$ and $comp(P) \nvDash_3 \exists S((Holdsat(rain,S), S < T) \vee (Holdsat(sprinkler\_on,S), S < T)$. In order to obtain that either is true that it rained or that the sprinkler was on, backwards reasoning needs to be employed, and more specifically closed world reasoning about rules (CWRr). With this term (CWRr) I will refer to the assumption that the only rules that hold about a certain φ in my representation of the world are the ones that I have. Hence, if φ is true this can only be the case because one of the antecedents that entail φ is the case. Symmetrically to the closed world reasoning about facts – according to which all information not explicitly true in the representation is assumed to be false–, the CWRr prescribes that all rules concerning φ not known by the agent nor present in the Knowledge Base are ignored to evaluate why φ holds[20].

The generation of the relevant hypotheses through CWRr can be easily modelled by integrity constraints in Abductive Logic Programming [65, 33]. An Abductive logic program is a triple $\langle P, Ab, IC\rangle$, where P is a finite set of (Horn) clauses as in definition A.9. Ab is a set of open or abducible predicates, i.e. predicates that are not defined in P and which cannot be derived. At least for a first introduction, I will follow the tradition of assuming that abducible do not appear as heads of clauses in P[21]. Finally, IC is a finite set of general clauses – i.e. disjunctions and existential quantifiers are allowed to appear in the head– called integrity constraints. Notice that even though integrity constraints will be formally written as normal conditional, to ease the explanation of their use I will use *IF* . . . , *THEN* . . . to represent an integrity constraint. From def. A.43 onwards, however, I will simply use $\varphi \rightarrow \psi$ to represent integrity constraints.

Intuitively, integrity constraints are better understood as prohibitions or obligations that the program P must satisfy. Alternatively, they can be thought of as non-truth functional, and as requesting action(s) to make the consequent true if the antecedent holds. Kowalski in Kowalski [65] uses the example of "IF it rains, THEN carry an umbrella" to introduce integrity constraints. While it may easily be true that it rains and I don't have an umbrella with me, this is still not sufficient to make the rule "If it rains, then carry an umbrella" false. Its intended meaning would be something akin to "In the case it rains, do something to have an umbrella".

More formally, an integrity constraint can be understood as a conditional

IF query $?\varphi$ succeeds, THEN query $?\psi$succeeds

[20] Following a convention in Logic Programming, I will say that a rule is *relevant*, *concerning* or *about* a proposition φ if φ is the head of the clause. In order to construct the completion of a program P, indeed, one takes all the clauses s.t. φ is its head and creates a disjunction of all the bodies $\psi_1, \ldots, \psi_n$. This formula $\varphi \leftrightarrow \bigvee_{i\in I} \psi_i$ is called the *definition* of φ. Furthermore, because clauses are written with the head in front – i.e. a rule of the form $\psi \rightarrow \varphi$ is usually written as $\varphi \leftarrow \psi$– it is visually easy to identify the clauses that say something about φ.

[21] This assumption is not strictly needed, but it simplifies the formalities. See Kowalski [65], p. 280.

Let P be a logic program, then φ may or may not hold wrt to P. Let P' be a n extension of P s.t. $T(P') \vDash_3 \varphi$, then $T(P') \vDash_3 \psi$.

The extension P' of P can be – in principle – found by looking at the SLDNF tree of $?\varphi$ and update P with the leaves of a finitely failed branch[22]. In order to appreciate how this works, the consequent of the integrity constraint is asked as the top query. Then the standard resolution rule is applied with clauses from the database for as long as possible. In general, this will terminate with a query $?G_i$ which cannot be further simplified (or with constraints). Normally, negation as failure would then be applied to the top query. Instead in the case of integrity constraints, the final query that couldn't be further simplified is used to update the program P, ensuring that the top query will succeed. In order to appreciate the informal sketch of abduction, it is necessary to introduce the definition of the resolution procedure for constraint logic programming. The definition, for simplicity's sake, only covers the Horn clauses case. For the more complex case describing resolution for goals with negations and existential quantifiers, the reader can have a look to definition A.36 in chapter A.3.3.

Definition A.42. (Resolution for CLP) Let $?c, b_1, \ldots, b_i, \ldots, b_n$ be a goal and $d_1, \ldots, d_k, c' \rightarrow a$ be a program clause, where c and c' are constraints, then a new goal $?c'', b_1, \ldots, d_1, \ldots, d_k, \ldots, b_n$ can be derived from these two clauses if c'', defined as $c'' = \{c \wedge \{b_i = a\} \wedge c'\}$, is satisfiable in $\mathscr{A}$.

Kowalski [65] uses integrity constraints to represent how our mental representations interact with the sensory stimuli and hence with "reality".For the present aims, it will be sufficient to use them to represent closed world reasoning (CWR) about rules, i.e. the idea that if there are n many rules about φ of the form

$$\psi_1 \rightarrow \varphi, \ldots, \psi_n \rightarrow \varphi$$

and φ , then it is allowed to infer that φ can holds only because one of the φ_i for $i \leq n$ holds.

An interesting application of integrity constraint can be found in the modelling of neuro-typical and autistic subjects in backward inference such as affirmation of the consequent and modus tollens[23].

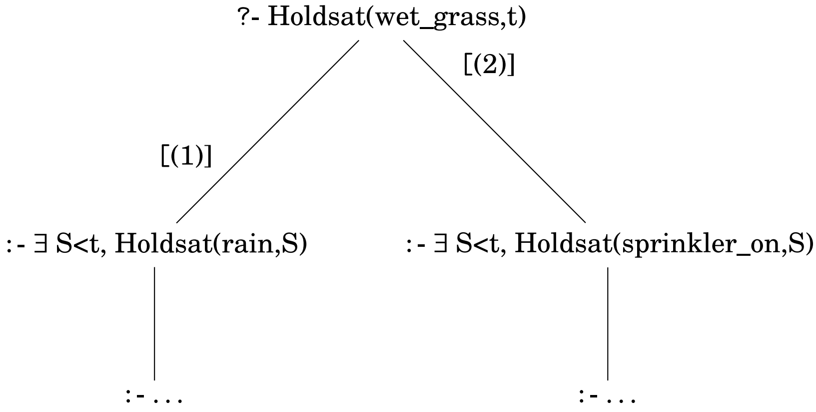
A.4.1 How abduction works

A good explanation for an observation should have certain properties such as being relevant, being minimal and being consistent with the rest of the available information. The relevance condition is immediately satisfied by reasoning backwards from the observation. Informally, indeed, the way to think of abduction is by looking at the SLD(NF) tree , where our observation is the initial query.

[22] [117, p. 198-193], [115, p. 103]

[23] See van Lambalgen and Stenning [117] (chapter 7), Smid [107],van Lambalgen and Smid [116] Pijnacker et al. [92].

Figure A.2: Abduction: an example



Let's consider again the example of the wet grass. If we have $?Holdsat(wet\_grass, t)$ denoting the proposition "the grass is green" as a query, then by definition of the resolution rule any step in a SLDNF tree has a resolvent iff the node and a clause share an atom. This ensures that only the he clauses that are related in the program are going to be involved in a backward chain. For the discussion of minimality and consistency, we refer the reader directly to Kowalski [65], pp. 139-141.

Intuitively, an abduction is successful if we obtain a good explanation for our observation(s) s.t. if we add this explanation to our program we are able to deduce what we observed. Formally, this means that a solution of a query Q is a set Δ of ground open predicates s.t.

1. Q holds wrt the program $P \cup \Delta$;

2. $P \cup \Delta$ satisfies the set of IC.

The precise meaning of "hold" and "satisfies" varies in the abundant literature about ALP: some authors formally make these terms precise in the sense of consistency, others in the sense of theoremhood, others on the basis of minimal models. For instance, according to the consistency view, an integrity constraint is satisfied if it is consistent with a program. On the theoremhood view instead, an integrity constraint is satisfied if is a consequence of the program in every model. In this work, I will follow Kowalski [65] and I adopt the minimal model view. According to it, a set of ground instances of the open predicates Ab is a solution of $?G$ iff $\{G\} \cup IC$ is true in some minimal model of $P \cup \Delta$. The minimal model for $P \cup \Delta$ in the case P is a Horn program it is simply its Herbrand model. The extensions for non-Horn programs and for negation are quite straight forward[24].

Definition A.43. (ALP Proof Procedure) Let G_0 be a query. The proof procedure uses forward and backward reasoning in an attempt to generate a solution Δ of G_0 by generating an abductive

[24]See Kowalski [65], chapter A6.

derivation $G_0, G_1, \ldots, G_n$ such that G_n contains the set Δ but no other goals that need to be solved. Each G_{i+1} is obtained from the previous G_i by one of the following inference rules:

1. (F1) Forward reasoning with a selected open atom A in G_i and an integrity constraint in IC.
   Suppose the integrity constraint has the form $A \wedge B \rightarrow C$ and G_i has the form $A \wedge G$. Then G_{i+1} is

   $$(B \rightarrow C) \wedge A \wedge G.$$ [25]

   (notice that this introduces a conditional into the goal clause. For this reason, we call the resulting goal clauses generalised goal clauses.)

2. (F2) Forward reasoning can also be use with a selected open atom A and a conditional in G_i.
   Suppose G_i has the form $(A \wedge B \rightarrow C) \wedge A \wedge G$, then G_{i+1} has the form

   $$(B \rightarrow C) \wedge A \wedge G$$

3. (B1) Backwards reasoning with a selected atom C in G_i and a clause in P
   Suppose the clause has the form $C \leftarrow D$ and G_i has the form $C \wedge G$, then G_{i+1} is

   $$D \wedge G$$

4. (B2) Backward reasoning with a selected atom C in a conditional in G_i having the form $(C \wedge B \rightarrow H) \wedge G$.
   Suppose $C \leftarrow D_1, \ldots, C \leftarrow D_m$ are all the clauses in P having conclusion C. Then G_{i+1} is

   $$(D_1 \wedge B \rightarrow H) \wedge \cdots \wedge (D_m \wedge B \rightarrow H) \wedge G$$

5. (F) Factoring between two copies of an open atom A in G_i
   If G_i has the form $A \wedge A \wedge G$ then G_{i+1} has the form $A \wedge G$. (Any previous application of F1 and F2 to any occurrence of A are deemed to have been done to the resulting single copy of A.)

6. (S) Splitting:
   If G_i has the form $(D_1 \vee \cdots \vee D_m) \wedge G$, then there are as many successor nodes G_{i+1} of the form $D_i \wedge G$ as there are disjuncts D_i. (Splitting needs to be performed when the conditions of an integrity constraint have been reduced to $true$, and the disjunctive conclusion has been conjoined to the sub-goals in G_i.

[25] Notice that this would be *IF* B, *THEN* C, and A and G. Henceforth, I will follow Kowalski [65] as represent integrity constraints as normal clauses (generally with $\rightarrow$ instead of $\leftarrow$).

7. (S) Replace $false \wedge C$ by $false$.

Definition A.44. (Successful abductive derivation) An abductive derivation $G_0, G_1, \ldots, G_n$ using these inference rules is a successfully terminating derivation of a set of open atoms Δ if and only if:

1. G_n is not false;
2. G_n has the form $(B_1 \to C_1) \wedge \cdots \wedge (B_m \to C_m) \wedge A_1 \wedge \ldots A_n$, $m \geq 0, n \geq 0$ where each A_i is an open atom;
3. no further applications of the inference rules can be performed on G_n, no matter which atom is selected;
4. $\Delta = \{A_1, \ldots, A_n\}$
5. the conditions B_i of the residues are not true in the minimal model of $P \cup \Delta$.

The residual conditionals $B_i \to C_i$ in a successfully terminating derivation are conditionals introduced by F1 but whose remaining conditions B_i are not true in the minimal model of $P \cup \Delta$. The conditions B_i of these residuals may consist solely of open atoms not in Δ; or they may contain closed atoms C that are not the conclusions of any clauses in P. In the latter case, it is as though there were a clause of the form $C \leftarrow false$ in P (as a result of which B_i is $false$, and the residual can be simplified to true and be ignored). The last condition of definition A.44 is needed to deal with the case of a successful derivation that does not terminates, i.e. one in which there is a Δ that makes true both the query the integrity constraint but the derivation loops. Checking that the conditions are not true requires recognising infinite failure, and this can be done through tabling[26] For Horn abductive programs and a ground Horn query, there is a soundness and completeness result that I merely state[27]. The soundess result refers to definition A.44 without the last condition.

Theorem 4. *(Soundess) Let $\langle P, Ab, IC \rangle$ be a Horn abductive logic program and G_0 a Horn ground query, then*

> *If there exists a successfully terminating derivation of Δ, then $\{G_0\} \cup IC$ is true in the minimal model of $P \cup \Delta$.*

The completeness results instead requires the last condition of definition A.44

Theorem 5. *(Completeness) Let $\langle P, Ab, IC \rangle$ be a Horn abductive logic program and G_0 a Horn ground query, then*

[26] Sagonas et al. [101].
[27] See Kowalski [65], chapter 6.

If $\{G_0\} \cup IC$ *is true in the minimal model of* $P \cup \Delta$*, then there exists a successful derivation of* Δ'*, such that* $\Delta' \subseteq \Delta$.

APPENDIX B

APPENDIX B: REALITY–PRETENCE REVISITED

On the basis of the theory for pretence developed across chapters 3 and 4, I hypothesise that children's understanding of pretend actions and events requires the understanding that pretended actions have different effects than real ones. An example of this, mentioned in chapter 2, is the following: if I work for 8 hours on my thesis, I expect certain effects to take place. Ideally, I would expect to have produced something intelligible to other people, or to type sentences that are grammatical and make some sense to me, or to think about some important problem and its solution. Even though at the end of the day I haven't produced anything materially observable, my actions (of typing or thinking or fixing) have certain effects. However, if I pretend to work for 8 hours, I wouldn't expect to have these effects. I may act as if I was typing, or as if I was thinking about some important problem, but I wouldn't expect to type grammatical and meaningful sentences nor to come across an insight about an important theoretical problem. I would say that whenever my pretending to work actually produces all and only the same effects as the real working[1], I am not pretending anymore.

The basic intuition behind this is that pretend actions have different effects from real actions and events. A child pretending that some pretend tea has been poured on her puppet's head, may pretend to dry the puppet's head or to wash the puppet, but – I believe – would be surprised if she found out that the puppet is actually wet.

A study on children's usage of the words "real","really" and "pretend" [11] suggests that they are used to express the authenticity of an object, more than its "ontological" status. This would explain the results in Baron-Cohen [4], Flavell et al. [29].

[1]As mentioned in chapter 3, the distinction between "real" consequences and "pretended" ones is best conceived as a spectrum of conditions. This allows to explain how children can behave really much as if their imaginary friend was around: it is sufficient that one small consequence of their pretended actions differs from the real consequence to mark a distinction between reality and pretence

However, in order to assess whether children understand pretend actions as having different effects than real actions, I propose to modify the pretence understanding task used in Kavanaugh and Harris [62]. Let me start by describing the paradigm used in the original experiment, to then introduce the intended modifications so as to test how children understand the effect of pretended actions.

B.1 The paradigm used in Kavanaugh and Harris (1994)

Kavanaugh and Harris [62] devised a non-verbal task to test whether children rely on gestures to guide and scaffold the comprehension of a pretend play – what they call a constructive slot-filling process. They further hypothesised that if children engage in such a "constructive process of visualisation", they would have been able to pick a picture representing the imaginary effect of the pretended transformation carried out by the experimenter. Therefore they designed a task in which children were asked to choose the picture depicting the imagined transformation among other pictures representing no change or irrelevant changes. The non-verbal nature of the task[2] allowed Kavanaugh and Harris to test whether children below 2 years are able to understand pretence transformation, but fail to engage in pretence because they don't know how to respond. Furthermore, the reduced linguistic complexity of the task permitted to test pretence comprehension for children with language difficulties, including autism.

I will now describe experiment 3 in Kavanaugh and Harris [62], which involved both children with autism and with mental retardation. The results can be compared with the ones of normally developing children, tested under the exact same conditions in experiment 2.

Procedure: the testing was preceded by a warm-up session, during which the children were shown two picture trios. The picture depicted familiar objects, and children were asked to point at each of them in succession. This was meant to familiarise children with both the experimenter and the task of pointing at one of the pictures.

The testing started with the experimenter saying "Now, let's play a game of pretence, shall we?". Children were given six test items – i.e. toy animals: duck, cat, bear, pig, rabbit and monkey. For each item, the experimenter placed it at approximately 30 cm from the child, with three pictures arranged from left to right between the item and the child. For expository convenience, let me describe the case in which the toy animal was a monkey. Of the three pictures, one pictured depicted the initial status of the toy animal (i.e. the monkey), one the effect of the imaginary transformation carried out by the experimenter (e.g. if the experimenter pretended to pour talcum on the monkey, the body of the monkey was all white), and one representing an irrelevant change (e.g. the monkey with red bars all over the body). After placing all these items, the experimenter

[2]The experimenter carried out six pretend actions modifying a chosen target, e.g. a puppet, and children where shown either two pictures – in experiment 1– or three pictures –experiments 2 and 3. Children were asked to indicate how the target looked after the pretend transformation (see Kavanaugh and Harris [62]). Hence, the experiment involved the linguistic request to indicate the imaginary outcome picture, but as reduced as possible,

said “See what I’m doing to the monkey!” and proceeded to enact the pretend transformation – in the monkey’ case taking the talcum can, shaking it on the monkey’s body. At this point, the experimenter asked “How does the monkey look like now? Does the monkey look like this picture now, or like this picture, or like this picture?” The pointing carried out by the experimenter always proceeded from left to right. If the child did not respond to this request, the experimenter re-phrased her request saying “Now, I’ve pretended to shake the talcum over the monkey. Does he look like this picture, or this one, or this one?”.

In the neuro-typical case, the subjects were 30 children ranging from 18 to 30 months (divided into two subgroup by age). In the autistic and mentally retarded case, the subjects were 24: 12 children with autism, and 12 with mental retardation. The mean mental age for the autistic group was 6 years and 6 months, while the mean chronological age was 9 years and 11 months. For the mentally retarded group, mean mental age was 6 years and 3 months, and chronological 9 years and 9 months.

The results displayed that children with autism performed better than chance, choosing mostly the correct picture, and in most errors pointing to the picture showing the animal as seen. The mentally retarded group, instead, performed at chance.

Finally, notice that all the transformations carried out by the experimenter in Kavanaugh and Harris [62] enter in the categories of functional play, since the props used were functionally appropriate to literally produce the intended affect. For instance, if the experimenter was pretending to pour tea on the toy duck, a tea-pot was used.

B.2 The proposed modification

In order to test whether normally developing children and with autism understand that pretend actions have *different* consequences from real actions, I prose to modify the experimental procedure in Kavanaugh and Harris [62] under two aspects:

1. instead of displaying functional plays, representing object substitutions;

2. instead of having one of the three pictures representing an irrelevant change in the target, depicting on it the *real* effects of the imaginary transformation, i.e. how the target would look like if the action was just as a normal action.

Let me give some examples of how I envisage the modified pretend acts. The ideal presentation of these stories is through a cartoon video displayed on a screen at 30 cm ca from the child, with three pictures disposed on the table between the child and the screen.

Example B.1. *A duck is sitting at a table, on which there is a banana. The duck picks up the banana, brings it to its ear and starts quacking into the banana.*
Three pictures are laid on the table: one represents the duck sitting at the table (no change), one represents the duck with a speech bubble coming out of its beak (imaginary transformation), and

one with two toy animals never displayed before (or two humans) talking into a banana phone (reality condition).

Example B.2. *A monkey is sitting straight on a chair. The monkey changes the way it's sitting on the chair and piggyback rides the chair, with its arms stretched in front of it, and emitting the sound "Vroom vroom".*
Three pictures are on the table: one represents the monkey sitting on the chair straight (no change), one depicts the monkey riding its chair around, one depicts a toy animal never displayed before (or a human) riding a chair on a street (reality condition).

Example B.3. *A pig is sitting in a kitchen, with a blue brick in front of it. The pig takes up the blue brick, brings it to its mouth and moves the mouth as to chew.*
Three pictures are on the table: one represents the monkey sitting with the blue brick in front of it, one depicting the pig with a visibly bigger belly and crumbs of blue brick around (intended change), and one with a never shown before animal (or human) with the blue brick in its/her mouth trying to chew it (reality condition).

Notice that because children with autism are sometimes observed trying to eat bricks in pretend play tasks (see Baron-Cohen [4]), the last type of trial would be particularly interesting so as to test whether the attempt to eat the bricks or other prompts used as food is due to a difficulty in comprehension, or to a difficulty in inhibiting the prepotent response – which would be due to an executive functions deficit.

Since object substitution is generally more difficult than functional play, both for normally developing children and the ones with autism, I would propose to carry out the experiment with a group of subjects of higher verbal and chronological age than the groups analysed in Kavanaugh and Harris [62].

BIBLIOGRAPHY

[1] D. R. Addis, L. Pan, M. Vu, N. Laiser, and D. L. Schacter. Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia*, 47:2222–2238, 2009.

[2] S. Baron-Cohem, A. Leslie, and U. Frith. Does the autistic child have a "Theory of Mind"? *Cognition*, 21:37–46, 1985.

[3] S. Baron-Cohen. Autism and symbolic play. *British journal of developmental psychology*, 5 (2):139–148, 1987.

[4] S. Baron-Cohen. Are autistic children "Behaviorists"? an examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders*, 19(4):579–600, 1989.

[5] S. R. Beck, D. J. Carroll, V. E. Brunsdon, and C. K. Gryg. Supporting children's counterfactual thinking with alternative modes of responding. *Journal of Experimental Child Psychology*, 108:190,Äì202, 2011.

[6] F. Berto. Impossible worlds and the logic of imagination. *Erkenntnis*, 2010.

[7] F. Berto and T. Schoonen. Conceivability and possibility: some dilemmas for humeans. *Synthese*, pages 1–19, 2017.

[8] S. Bigham. Comprehension of pretence in children with autism. *British Journal of Developmental Psychology*, 26(2):265–280, 2008.

[9] S. Bigham. Impaired competence for pretense in children with autism: Exploring potential cognitive predictors. *Journal of Autism and Developmental Disorders*, 40:30,Äì38, 2010.

[10] G. Boolos, R. Jeffrey, and J. B. Burgess. *Computability and Logic*. Cambridge University Press, Cambridge, 1989.

[11] L. Bunce and M. Harris. "I saw the real Father Christmas!" children's everyday uses of the words *real*, *really* and *pretend*. *British Journal of Developmental Psychology*, 26:445–55, 2008.

[12] G. M. Burghardt. Defining and recognizing play. In P. E. Nathan, editor, *The Oxford Handbook of the Development of Play*. Oxford University Press, Oxford, 2011.

[13] R. Byrne. *The Rational Imagination*. MIT Press, Cambridge, MA, 2005.

[14] P. Carruthers. Why pretend? *The architecture of the imagination*, pages 89–109, 2006.

[15] P. Carruthers and P. K. Smith. *Theories of theories of mind*. Cambridge University Press, 1996.

[16] D. Chalmers. Does conceivability entail possibility? *Conceivability and possibility*, pages 145–200, 2002.

[17] D. J. Chalmers. The tyranny of the subjunctive. *na*, 1998.

[18] T. Charman and S. Baron-Cohen. Brief report: Prompted pretend play in autism. *Journal of autism and developmental disorders*, 27(3):325–332, 1997.

[19] J. Craig and S. Baron-Cohen. Creativity and imagination in autism and asperger syndrome. *Journal of autism and developmental disorders*, 29(4):319–326, 1999.

[20] R. Cristi-Vargas, J. Perner, and E. Rafetseder. Counterfactual reasoning: Developing a sense of "Nearest Possible World". *Child Development*, 81(1):376–389, 2010.

[21] G. Currie and I. Ravenscroft. *Recreative minds: Imagination in philosophy and psychology*. Oxford University Press, 2002.

[22] G. O. Deák, A. D. Pick, and S. D. Ray. Matching and naming objects by shape or function: Age and context effects in preschool children. *Developmental Psychology*, 38:503–518, 2002.

[23] M. Dias and P. L. Harris. The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology*, 6:207–21, 1988.

[24] M. Dias and P. L. Harris. The influence of the imagination on reasoning by young children. *British Journal of Developmental Psychology*, 8(4):305–318, 1990.

[25] T. Doggett and A. Egan. Wanting things you don't want: The case for an imaginative analogue of desire,. *Philosophers' Imprint*, 7:1–17, 2007.

[26] DSM-IV. *Diagnostic and statistical manual of mental disorders*. American Psychiatric Association, Washington, DC, 1994.

[27] S. Feferman. Toward useful type-free theories. i. *The Journal of Symbolic Logic*, 49(1): 75–111, 1984.

[28] M. Fitting. Bilattices and the semantics of logic programming. *Journal of Logic Programming*, 11:91‚Äì116, 1991.

[29] J. H. Flavell, E. R. Flavell, and F. L. Green. Development of the appearance-reality distinction. *Cognitive psychology*, 15(1):95–120, 1983.

[30] W. Friedman. *About time: Inventing the fourth dimension.* The MIT press, 1990.

[31] C. Frith. What do imaging studies tell us about the neural basis of autism? In *Autism: Neural Basis and Treatment Possibilities*, pages 149–166. Novartis Foundation Symposium, Wiley, New York, 2003.

[32] C. Frith. The self in action: lessons from delusions of control. *Consciousness and cognition*, 14(4):752–770, 2005.

[33] T. H. Fung and R. Kowalski. The iff proof procedure for abductive logic programming. *The journal of Logic Programming*, 33(2):151–165, 1997.

[34] V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti. Action recognition in the premotor cortex. *Brain*, 119(2):593–609, 1996.

[35] T. Gendler. Imagination. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.

[36] T. P. German and S. Nichols. Children's counterfactual inferences about long and short causal chains. *Developmental Science*, 6(5):514–523, 2003.

[37] A. I. Goldman. *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press, 2006.

[38] R. M. Gordon. The simulation theory: Objections and misconceptions. *Mind & Language*, 7 (1-2):11–34, 1992.

[39] G. Harman. *Change in view: Principles of reasoning.* Cambridge University Press, 2008.

[40] P. L. Harris. *Children and Emotion*. Blackwell, Oxford, 1989.

[41] P. L. Harris. From simulation to folk psychology: The case for development. *Mind and Language*, 7(1-2):120– 145, 1992.

[42] P. L. Harris. Pretending and planning. In S. Baron-Coehn, H. Tager-Flusberg, and D. J. Cohen, editors, *Understanding other minds: perspectives from autism*. Oxford University Press, Oxford, 1993.

[43] P. L. Harris. *The Work of the Imagination: understanding children's worlds*. Blackwell, Oxford, 2000.

[44] P. L. Harris, R. D. Kavanaugh, H. M. Wellman, and A. K. Hickling. Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, 58(1):1–107, 1993.

[45] P. L. Harris, T. German, and P. Mills. Children's use of counterfactual thinking in causal reasoning. *Cognition*, 61:233–259, 1996.

[46] P. L. Harris, R. D. Kavanaugh, and L. Dowson. The depiction of imaginary transformations: Early comprehension of a symbolic function. *Cognitive Development*, 12(1):1–19, 1997.

[47] P. L. Harris, E. S. Pasquini, S. Duke, J. J. Asscher, and F. Pons. Germs and angels: the role of testimony in young children's ontology. *Developmental Science*, 9(1):76–96, 2006.

[48] J. Heal. Replication and functionalism. In B. J, editor, *Language, Mind, and Logic*. Cambridge University Press, Cambridge, MA, 1986.

[49] R. P. Hobson. *Autism and the development of mind*. Psychology Press, 1993.

[50] W. Hodges. *Model theory*. Cambridge University Press, Cambridge, 1993.

[51] C. Hughes and J. Russell. Autistic children's difficulty with mental disengagement from an object: Its implications for theories of autism. *Developmental psychology*, 29(3):498, 1993.

[52] D. Hume. *Treatise of Human Nature*. Clarendon Press, Oxford, 1968.

[53] P. R. Huttenlocher. Dendritic and synaptic development in human cerebral cortex: time course and critical periods. *Developmental Neuropsychology*, 16(3):347–349, 1999.

[54] M. B. Jackson. On the epistemic value of imagining, supposing and conceiving. In *Knowledge Through Imagination*, pages 41–60. Oxford University Press, Oxford, 2016.

[55] G. . Jäger and R. F. Stärk. Chapter ix a proof-theoretic framework for logic programming. In S. R. Buss, editor, *Handbook of Proof Theory*. Elsevier Science B.V., 1998.

[56] C. Jarrold. A review of research into pretend play in autism. *Autism*, 7(4):379–90, 2003.

[57] C. Jarrold and C. Conn. The development of pretend play in autism. In *The Oxford Handbook of the Development of Play*. Oxford University Press, Oxford, 2011.

[58] C. Jarrold, P. Smith, J. Boucher, and P. Harris. Comprehension of pretense in children with autism. *Journal of Autism and Developmental Disorders*, 24:433–55, 1994.

[59] C. Jarrols, J. Boucher, and P. K. Smith. Generativity deficits in pretend play in autism. *British journal of Developmental Psychology*, 14(3):275–30, 1996.

[60] J.Hawkins, R. D. Pea, J. Glick, and S. Scribner. " merds that laugh don't like mushrooms": Evidence for deductive reasoning by preschoolers. *Developmental Psychology*, 20(4):584, 1984.

[61] R. Kavanaugh. Origins and consequences of social pretend play. In P. E. Nathan, editor, *The Oxford Handbook of the Development of Play*. Oxford University Press, Oxford, 2011.

[62] R. D. Kavanaugh and P. L. Harris. Imagining the outcome of pretense transformations: assessing the competence of normal children and children with autism. *Developmental Psychology*, 30(6):846–854, 1994.

[63] A. Kind and P. Kung. Introduction: the puzzle of imaginative use. In *Knowledge Through Imagination*, pages 1–39. Oxford University Press, Oxford, 2016.

[64] A. Kind and P. Kung. *Knowledge Through Imagination*. Oxford University Press, Oxford, 2016.

[65] R. Kowalski. *Computational Logic and Human Thinking: how to be artificially intelligent*. Cambridge University Press, Cambridge, 2011.

[66] R. Kowalski and M. Sergot. A logic-based calculus of events. *New Generation Computing*, 4:67–95, 1986.

[67] S. A. Kripke. Identity and necessity. In *Identity and Individuation*, pages 135–164. New York University Press, New York, 1971.

[68] B. Landau, L. B. Smith, and S. S. Jones. The importance of shape in early indexical learning. *Cognitive Development*, 3:299–321, 1988.

[69] P. Langland-Hassan. On choosing what to imagine. In A. Kind and P. Kung, editors, *Knowledge Through Imagination*, pages 61–84. Oxford University Press, Oxford, 2016.

[70] H. J. Leevers and P. Harris. Counterfactual syllogistic reasoning in normal 4-year-olds, children with learning disabilities, and children with autism. *Journal of Experimental Child Psychology*, 76:64‚Äì87, 2000.

[71] A. M. Leslie. Pretense and representation: The origins of" theory of mind.". *Psychological review*, 94(4):412, 1987.

[72] A. M. Leslie. Pretending and believing: Issues in the theory of tomm. *Cognition*, 50(1): 211–238, 1994.

[73] D. Lewis. *Counterfactuals. Malden*. MA: Basil Blackwell, Ltd, 1973.

[74] V. Lewis and J. Boucher. Spontaneous, instructed and elicited play in relatively able autistic children. *British Journal of Developmental Psychology*, 6(4):325–339, 1988.

[75] A. Lillard, A. M. Pinkham, and E. Smith. Pretend play and cognitive development. In U. Goswami, editor, *The Wiley-Blackwell handbook of childhood cognitive development*, pages 285–311. Wiley-Blackwell Chichester, 2011.

[76] A. S. Lillard. Pretend play skills and the child's theory of mind. *Child development*, 64(2): 348–371, 1993.

[77] A. S. Lillard. Mother‚Äìchild fantasy play. In P. E. Nathan, editor, *The Oxford Handbook of the Development of Play*. Oxford University Press, Oxford, 2011.

[78] J. W. Lloyd. *Foundations of Logic Programming*. Springer-Verlag, Berlin, 1984.

[79] J. L. Mackie. *The Cement of the Universe: a study of causation*. Oxford University Press, Oxford, 1974.

[80] A. Martelli and U. Montanari. An efficient unification algorithm. *ACM Transactions on Programming Languages and Systems*, 4:258‚Äì282, 1982.

[81] C. McGinn. *Mindsight: Image, Dream, Meaning*. Harvard University Press, Cambridge, MA, 2004.

[82] R. Miller and M. Shanahan. *The event calculus in classical logic-alternative axiomatisations*. Linköping University Electronic Press, 1999.

[83] M. Myles-Worsley, C. C. Cromer, and D. H. Dodd. Children's preschool script reconstruction: Reliance on general knowledge as memory fades. *Developmental Psychology*, 22(1):22–30, 1986.

[84] K. Nelson. Event representations, narrative development and internal working models. *Attachment & Human Development*, 1(3):239–252, 1999.

[85] S. Nichols and S. Stich. *Mindreading. An integrated account of pretence, self-awareness and understanding other minds*. Oxford University Press, Oxford, 2003.

[86] A. Paivio. *Mental representations: A dual coding approach*. Oxford University Press, New York, 1986.

[87] J. Perner, S. R. Leekam, and H. Wimmer. Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5(2):125–137, 1987.

[88] J. Perner, E. Rafetseder, and M. Schwitalla. Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, 114:389‚Äì404, 2013.

[89] D. M. Peterson and D.M.Bowler. Counterfactual reasoning and false belief understanding in children with autism. *Autism: the international Journal of Research and Practice*, 4(4): 391–405, 2000.

[90] D. M. Peterson and K. J. Riggs. Adaptive modelling and mindreading. *Mind and Language*, 14(1):80–112, 1999.

[91] J. Piaget. *Play, dreams and imitation in childhood*. Routledge, 1962.

[92] J. Pijnacker, B. Geurts, M. van Lambalgen, C. C. Kan, J. K. Buitelaar, and P. Hagoort. Defeasible reasoning in high-functioning adults with autism: Evidence for impaired exception-handling. *Neuropsychologia*, 47:644‚Äì651, 2009.

[93] M. Prior and W. Hoffmann. Brief report: Neuropsychological testing of autistic children through an exploration with frontal lobe tests. *Journal of autism and developmental disorders*, 20(4):581–590, 1990.

[94] C. D. P. K. Ramli. Logic programs and three-valued consequence operators, 2009.

[95] C. A. Richards and J. A. Sanderson. The role of imagination in facilitating deductive reasoning in 2-, 3-and 4-year-olds. *Cognition*, 72(2):B1–B9, 1999.

[96] K. J. Riggs, D. Peterson, E. J. Robinson, and P. Mitchell. Are errors in false belief task symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, 13: 73–90, 1998.

[97] G. Rizzolatti and L. Craighero. The mirror-neuron system. *Annu. Rev. Neurosci.*, 27: 169–192, 2004.

[98] N. J. Roese and J. M. Olson. *What might have been: the social psychology of counterfactual thinking*. Erlbaum, Mahwah, NJ, 1995.

[99] J. E. Russell. *Autism as an executive disorder.* Oxford University Press, 1997.

[100] G. Ryle. *The concept of mind (1949)*. Hutchinson, London, 1984.

[101] K. Sagonas, T. Swift, and D. Warren. Xsb as an efficient deductive database. *SIGMOD Record*, 23(2):442–453, 1994.

[102] D. L. Schacter, D. R. Addis, and R. L. Buckner. Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, 8:657–661, 2007.

[103] F. J. Scott and S. Baron-Cohen. Logical, analogical, and psychological reasoning in autism: A test of the cosmides theory. *Development and Psychopathology*, 8(01):235–245, 1996.

[104] F. J. Scott, S. Baron-Cohen, and A. Leslie. If pigs could fly: A test of counterfactual reasoning and pretence in children with autism. *British Journal of Developmental Psychology*, 17: 349–362, 1999.

[105] M. Shanahan. Representing continuous change in the event calculus. *ECAI*, 90:598–603, 1990.

[106] K. Shanton and A. Goldman. Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(4):527–538, 2010.

[107] H. Smid. Reasoning with rules and exceptions in autism, 2005.

[108] R. Stärk. *From logic programs to inductive definitions*. na, 1996.

[109] P. F. Strawson. Imagination and perception. In *Experience and Theory*, pages 31–54. University of Massachusetts Press, Amherst, MA, 1970.

[110] P. J. Stuckley. Negation and constraint logic programming. *Information and Computation*, 118(1):12–33, 1995.

[111] N. J. Thomas. Mental imagery. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2017 edition, 2017.

[112] A. Treisman. Focused attention in the perception and retreival of multidimensional stimuli. *Cognition and Psychopsysics*, 22:1–11, 1977.

[113] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[114] A. Treisman and H. Schmidt. Illusory conjunctions in the perception of objects. *Cognitive psychology*, 14:107–141, 1982.

[115] M. van Lambalgen and F. Hamm. *The Proper Treatment of Events*. Blackwell, Oxford, 2004.

[116] M. van Lambalgen and H. Smid. Reasoning patterns in autism: rules and exceptions. In *Proc. Eighth International Colloquium on Cognitive Science Donostia/San Sebastian. Kluwer*. Citeseer, 2004.

[117] M. van Lambalgen and K. Stenning. *Human Reasoning and Cognitive Science*. MIT press, Cambridge, MA, 2012.

[118] N. van Leeuwen. Imagination is where the action is. *The Journal of Philosophy*, 108:55–77, 2011.

[119] L. S. Vygotsky. Imagination and creativity in childhood. *Journal of Russian and East European Psychology*, 42(1):7–97, 2004.

[120] K. Walton. *Mimesis as Make-Believe*. Harvard University Press, Cambridge, MA, 1990.

[121] D. S. Weisberg. Pretend play. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(3): 249–261, 2015.

[122] T. Williamson. *The Philosophy of Philosophy*. Blackwell, Oxford, 2007.

[123] T. Williamson. Knowing by imagining. In *Knowledge Through Imagination*, pages 113–123. Oxford University Press, Oxford, 2016.

[124] H. Wimmer and J. Perner. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1):103–128, 1983.

[125] S. Yablo. Is conceivability a guide to possibility? *Philosophy and Phenomenological Research*, 53(1):1–42, 1993.