Children’s cognition of continuous causal processes, the role of spatial-temporal analysis above/beyond other candidate predictors

Selma Dündar-Coecke

UCL

Thesis submitted for the degree of Doctor of Philosophy

2019
I hereby declare that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature .................................................................

Word count (exclusive of abstract, impact statement, dissemination, acknowledgements, appendices, table of contents, reference list and lists of figures and tables):

70.941 words.
Abstract

Causal reasoning is a fundamental component of any credible form of human thinking and has been a core topic of various fields, for the reason that understanding the means by which people grasp physical causality is crucial to the formation of everyday and scientific thinking. Causal relations can be conceived from discrete events and also from temporally continuous processes. However, the process view of causation is highly neglected in psychology literature. This capacity however paves the way for understanding various fundamental issues ranging from i.e. food security to climate change. How or when the ability to reason about continuous causal processes evolves into a mature form is unknown. What kind of cognitive abilities are involved in this evolution is largely unexplored. This thesis aims to provide the first systematic investigation on these. The thesis contains five chapters. The first chapter has focused on the rationale and explained the paradigms and methodological approaches used across three studies. The second chapter has examined whether a single datum suffices to establish a causal connection, and whether children’s understanding varies depending on the way a causal phenomenon is presented. This target has been extended in the third chapter by considering the developmental trajectories: how children move from stage to stage; how various cognitive competences play a role in these. This chapter has also concerned whether most promising candidate abilities, such as spatial, spatial-temporal, statistical thinking indexed by probability and covariation analysis predict children’s thinking of causal processes above/beyond verbal and nonverbal abilities. The fourth chapter has evaluated the replicability of the results and provided further evidence on the candidate predictors by extending the socioeconomic background of the sample. These three chapters have included their dedicated discussions. However, the fifth chapter has rendered a holistic picture and provided a detailed discussion on the meaning of the outcomes for the research questions.
Impact Statement

Of course society needs a STEM (Science, Technology, Engineering, Mathematics) research workforce, but it more crucially needs successful reasoners in their causal judgments for full participation in a globally shaped enhanced life. Given that science literacy is increasingly linked to economic growth and is necessary for finding solutions to complex social, biological, and environmental issues, a population level engagement is needed to deal with demands of 21st century life.

Not only knowledge, but also understanding of science and its methodologies are fundamental. This thesis focuses on this fundamental need, and aims to take an important step by providing robust evidence for unexplored cognitive abilities directing early scientific thinking: understanding the nature of which is necessary for teaching and learning. The way to develop this particular thinking is not through transmission of the content, instead by developing causal thinking, in particular mechanism level understanding. At the core of this, the ability to extract spatial-temporal information to understand about causal processes seems to be a key element, highly relevant to effective observation. Apparently, this is accompanied by the ability to construct mental transformations/imagery to infer invisible factors; potentially by probabilistic thinking to deal with unseen exceptions; increase in both daily and scientific vocabulary; and also development of other nonverbal abilities, as this thesis delineated. These elements can guide (1) why the gap between “literate” and ‘scientifically literate’ may start to widen in the population, and (2) how this gap can be closed in early years by effective interventions.

The outcomes therefore are oriented to both academic and non-academic users, and have the potential to substantially inform developmental, cognitive and educational science, alongside the development of pedagogical approaches targeting its users, such as children, teachers, their parents, producers of learning materials, school communities, and relevant institutions.
Acknowledgements

I would like to thank to the Economic and Social Research Council for funding this project, and as part of this, UCL, Bloomsbury and East London ESRC-DTP team for their support. I am thankful to the children, teachers who participated in the studies, and also the head teachers who allowed me to work for extended periods of time within their schools. I should in particular thank to two teachers, Jim Brown and Kate Jeager, from two primary schools located in Oxford, and Thames Valley Science Network coordinator for their effort and support on the applicability of the ideas in school settings.

I would like to thank to Professor Andrew Tolmie for his continuous support, advise, and encouragements from the very beginning of this project; to Dr. Anne Schlottmann for her advise, and sharing her experience; to Professors Emily Farran and Jane Hurry for taking part in my upgrade panel and making comments; to Professors David Lagnado and Teresa McCormack for examining my thesis.

I would also thank to Dina Khatun for running one of the subsidiary investigations as part of her thesis under our supervision; Yu Kee, Janice Howard for taking further some of the findings generated by this thesis as part of their masters program under our supervision.

I would like to extend my thanks to my family and friends, in particular my daughter and husband, for understanding and being supportive at all stages.
Chapter 3

The way of information presentation may affect children’s ability to predict,
describe and explain continuous causal processes

2.1 Introduction ................................................................. 65
2.2 The role of familiarity in accessing causal knowledge ......................... 66
2.3 Effects of information presentation ......................................... 70
2.4 The present study ................................................................ 73
2.5 Method ................................................................. 76
  2.5.1 Design .................................................................. 76
  2.5.2 Participants .......................................................... 76
  2.5.3 Materials and Procedure .......................................... 77
    2.5.3.1 Causal tasks .................................................... 77
      2.5.3.1.1 Condition 1 (paired-contrasting instances) .... 78
      2.5.3.1.2 Condition 2 (sequential instances) .......... 81
      2.5.3.1.3 Condition 3 (jumbled instances) ............ 83
      2.5.3.1.4 Scoring system ....................................... 85
    2.5.3.2 Measures of verbal and nonverbal ability .......... 88
2.6 Results ................................................................ ........ 88
  2.6.1 Comparability of conditions ............................................ 88
  2.6.2 Effects of condition on causal task performance ................ 88
  2.6.3 Relationships between causal indices and verbal/nonverbal ability .... 94
2.7 Discussion .................................................................... 96
  2.7.1 Causal inference from observation .............................. 97
  2.7.2 Effects of comparison and information presentation ............ 98
2.8 Conclusion .................................................................... 100

Chapter 3........................................................................ 102
Children's reasoning about continuous causal process: the role of verbal, nonverbal, spatial, spatial-temporal, and statistical analysis

3.1 Introduction .......................................................................................................................... 102
3.2 The role of spatial and spatial-temporal cognition in causal thinking ....................... 102
  3.2.1 Causal thinking and the role of spatial-temporal cognition ................................. 104
  3.2.2 Causal thinking and the role of spatial ability .................................................. 109
3.3 Causal thinking and the role of statistical inference ...................................................... 112
  3.3.1 How can probabilistic thinking play a role in causal processes? .................... 113
  3.3.2 How can covariation information play a role in causal processes? ............... 119
3.4 Method .................................................................................................................................. 122
  3.4.1 Design ............................................................................................................................ 122
  3.4.2 Participants ...................................................................................................................... 123
  3.4.3 Materials and procedure .............................................................................................. 123
    3.4.3.1 Measures of verbal and nonverbal abilities .................................................... 123
    3.4.3.2 Causal tasks ............................................................................................................ 124
    3.4.3.3 Measures of spatial-temporal analysis ............................................................. 129
      3.4.3.3.1 The flow of liquid task .................................................................................. 129
      3.4.3.3.2 The speed task ............................................................................................... 131
    3.4.3.4 Measures of spatial analysis ................................................................................. 133
      3.4.3.4.1 The monkey mental rotation ....................................................................... 133
      3.4.3.4.2 Paper-folding ................................................................................................. 134
    3.4.3.5 Statistical thinking tasks ....................................................................................... 135
3.5 Results .................................................................................................................................... 139
  3.5.1 Developmental trajectories ......................................................................................... 139
    3.5.1.1 Causal tasks .......................................................................................................... 139
    3.5.1.2 Spatial-temporal measures .................................................................................. 144
    3.5.1.3 Verbal, nonverbal and spatial ability .................................................................... 146
    3.5.1.4 Statistical tasks ..................................................................................................... 147
  3.5.2 What predicts children's reasoning of causal processes? ........................................ 148
    3.5.2.1 Correlations between variables ......................................................................... 148
    3.5.2.2 Hierarchical regression models .......................................................................... 150
    3.5.2.3 Nature of shared variances between predictors ................................................. 154
    3.5.2.4 Path analysis ........................................................................................................ 155
  3.5.3 Identification of causal mechanisms ........................................................................... 156
3.6 Discussion ............................................................................................................................. 157
  3.6.1 The development of causal reasoning about continuous processes ..................... 157
  3.6.2 Do verbal and nonverbal abilities predict children's reasoning about continuous causal processes? .......................................................... 160
  3.6.3 The development of spatial-temporal analysis ......................................................... 161
  3.6.4 What does predict children's reasoning of continuous causal processes? .......... 165

Chapter 4..................................................................................................................................... 167

Replication: the role of verbal, nonverbal, spatial, spatial-temporal, and statistical information to infer continuous causal processes across development ......................................................................................................................... 167

4.1 The rationale of the replication study .............................................................................. 167
4.2 Method .................................................................................................................................. 173
  4.2.1 Participants ...................................................................................................................... 173
  4.2.2 Design, Materials and Procedure ............................................................................... 173
    4.2.2.1 Causal tasks ........................................................................................................... 174
    4.2.2.2 Measures of spatial-temporal analysis ............................................................... 179
    4.2.2.3 Measure of spatial ability ...................................................................................... 184
4.2.2.4 Measures of statistical ability ................................................................. 184

4.3 Results ............................................................................................................. 186
  4.3.1 Developmental trajectories ...................................................................... 186
  4.3.1.1 Causal tasks ......................................................................................... 186
  4.3.1.2 Spatial-temporal measures ................................................................. 188
  4.3.1.3 Verbal, nonverbal, and spatial ability ................................................. 192
  4.3.1.4 Statistical tasks .................................................................................... 193
  4.3.2 What does predict children’s reasoning about causal processes? .......... 194
  4.3.2.1 Correlations between variables .......................................................... 194
  4.3.2.2 Hierarchical regression models ......................................................... 195
  4.3.2.3 Path analysis ....................................................................................... 201
  4.3.2.4 Nature of shared variances between predictors ................................. 202
  4.3.3 Identification of causal mechanisms ...................................................... 204
  4.3.4 Significance of correlations in two unequal samples in chapter three and four
                                                                                   ........................................................................................................... 205

4.4 Discussion ........................................................................................................ 206

4.5 General discussion .......................................................................................... 208
  4.5.1 Developmental trajectories .................................................................. 208
  4.5.2 What predicts children’s reasoning about continuous causal processes... 213
  4.5.2.1 Spatial-temporal measures .................................................................. 213
  4.5.2.2 Spatial measures ............................................................................... 217
  4.5.2.3 Verbal and nonverbal measures ....................................................... 218
  4.5.2.4 Nature of shared variances between predictors ............................... 219
  4.5.3 The cognitive processes involved in children’s spatial-temporal analysis ................................................................................................................. 222

Chapter 5 ............................................................................................................... 224

Overall discussion .................................................................................................. 224
  5.1 Children’s explanations progress differently in specific domains ............ 224
  5.2 Can children grasp of causal mechanism from observation? ................. 226
  5.3 How far verbal and nonverbal abilities important for causal thinking? .... 228
  5.4 On the nature of spatial-temporal analysis .............................................. 231
  5.5 Can children understand causal mechanism from single observations? ... 234
  5.6 The meaning of these outcomes for education ........................................ 235

References ............................................................................................................. 237

Appendix 1 ............................................................................................................. 275
List of Tables

Table 1. Scoring system for causal tasks .............................................................. 86
Table 2. Mean score (sd) by age group and condition on total causal measure 
(max=21), prediction, description (max=6), inference (max=9), mechanism 
(max=3), initial inference (max=9; conditions 2 and 3 only), vocabulary (max=43), 
block design (max=58) ...................................................................................... 90
Table 3. Partial correlations between causal measures, controlling for age (significant 
associations in bold) ......................................................................................... 92
Table 4. Frequency of children making different numbers of mechanism level inferences 
by level of description ....................................................................................... 94
Table 5. Hierarchical regression analysis with overall causal measure score as dependent 
variable (significant predictors in bold) .......................................................... 95
Table 6. A typical 2x2 contingency table, with cause and effect as the two variables. 120
Table 7. Scoring system for causal tasks .............................................................. 127
Table 8. Examples of explanation responses ...................................................... 128
Table 9. Mean score (sd) by age group on total causal score (max=21), prior knowledge, 
description (max=6), explanation (max=9), and mechanism (max=3) ............. 143
Table 10. Mean score (sd) by age group on rotation (max=16), paper folding (max=14), 
vocabulary, and block design .......................................................................... 146
Table 11. Mean score (sd) on randomness (max=2), marbles, cards (max=10), 
covariation (max=12) ...................................................................................... 148
Table 12. Zero-order and partial correlations (zero-order above diagonal, N=107; partial 
correlations below diagonal, controlling for age in months, verbal and nonverbal 
ability ............................................................................................................... 149
Table 13. Hierarchical regression with total causal score as dependent variable 
(significant predictors in bold) .......................................................................... 151
Table 14. Hierarchical regression with components as dependent variables (significant 
predictors in bold) ......................................................................................... 152
Table 15. Rotated component matrix of the four-factor model ............................ 154
Table 16. Percentages of children obtaining perfect scores on the flow of liquid task at 
each level of explanation response ................................................................ 157
Table 17. Scoring system for causal task ............................................................ 177
Table 18. Level 3 and 4 explanation responses ................................................... 178
Table 19. Mean score (sd) by age group on total score (max=33), description (max=3), 
prediction, justification (max=9), explanation (max=12) and mechanism (max=3) 187
Table 20. Zero order and partial correlations between spatial-temporal measures 
(significant associations in bold) ................................................................. 191
Table 21. Mean score (sd) by age group on rotation (max=16), vocabulary and block 
design ............................................................................................................. 192
Table 22. Zero-order and partial correlations between measures (significant associations 
in bold) ......................................................................................................... 194
Table 23. Hierarchical regression with total causal score as dependent variable 
(significant predictors in bold) ......................................................................... 196
Table 24. Hierarchical regression with components as dependent variables (significant 
predictors in bold) ..................................................................................... 198
Table 25. Three factor solution for the relationship between exponential flow of liquid, DTV, log block design, marbles, and covariation

Table 26. Percentages of children obtaining perfect scores on the flow of liquid task at each level of explanation response

Table 27. Mean DTV scores (sd) at each level of inference response

Table 28. Mean score (sd) by age group on VS1 (max=13), VS2 (max=8), and AS (max=3)

Table 29. Zero-order and partial correlations
List of Figures

Figure 1. The characteristics of processes (obtained from Sowa, 2000) ......................... 14
Figure 2. Materials used in the causal tasks ...................................................................... 78
Figure 3. Scores by age groups and condition on total causal performance (max=21), prediction, description (max=6), inference (max=9), and initial inference (max=9; conditions 2 and 3 only) .................................................................................................................. 89
Figure 4. Flow of liquid task ........................................................................................... 129
Figure 5. Example configuration of the bunnies at the start of a speed task trial ............ 132
Figure 6. Spatial tasks: (a) monkey mental rotation, (b) paper folding task ................. 134
Figure 7. (a) The four trays for the four trials of the Marbles task, in trial order; (b) the deck shown in the first trial of the cards task - only one smiley and one frown were dealt face up, the other two cards were shown face down .............................................. 136
Figure 8. Practice display of the covariation task .......................................................... 137
Figure 9. First display of covariation task (perfect covariation) ...................................... 138
Figure 10. Profile of scores on (a) sinking, (b) absorption, (c) solution (max=7) .......... 140
Figure 11. Profile of scores on (a) prior knowledge (max=6), (b) description (max=6), causal explanation (max=9) .................................................................................................................. 142
Figure 12. Profile of scores on (a) flow of liquid (max=10), and (b) speed (max=17) 145
Figure 13. Developmental trajectories .............................................................................. 147
Figure 14. Extended mediation model for the effects of log block design, covariation, marbles and flow of liquid on causal total score (subsidiary effects in grey) .......... 155
Figure 15. Materials in the correct order of prediction for the causal measures ............ 175
Figure 16. Materials for the distance-time-velocity and rotation tasks ............................ 182
Figure 17. One of the four trials of the covariation task employed in Study 2, displaying the first deck with 75% co-occurrence between starts and circles ................................................. 185
Figure 18. Profile of scores on (a) flow of liquid (max=12), (b) distance, (c) time, and (d) velocity (max=6 for all) ........................................................................................................... 190
Figure 19. Developmental trajectories .............................................................................. 193
Figure 20. Path model including standardised coefficients for the effects of exponential flow of liquid, DTV, log block design, age and vocabulary on causal total score 201
Figure 21. Further extended mediation model for the effects of log block design, covariation, marbles and flow of liquid on overall causal performance ....................... 203
Figure 22. Example configurations of bunnies at the start of ........................................... 275
Chapter 1
Introduction to the thesis

1.1 Overview

The term ‘information’ refers to anything that conveys meaning, and cannot be reduced to any specific source or representation (Floridi, 2011). Mostly causal relations ‘carry’ the information that is perceived, stored, discussed, used, or changed by agents (e.g. Sloman, 2005). Even primitive humans began to realize that certain things regularly cause other things, hence interest in causation is longstanding (Pearl & Mackenzie, 2018), evolving from ancient philosophy, mostly explained by Aristotle’s four causes, where the focus was on the content of living mechanisms and metaphysics, to Medieval science, atomists, Descartes, Newton, LaPlace, Hume, early 20th Century theories, and later quantum mechanics. In a broad sense, the message from history is clear: humans can extract information from various sources directly or indirectly to infer cause-effect relationships, and this can impact crucial aspects of our lives, from psychological and social to economic.

Causality is a hot topic, and various approaches and models explain how people reason about causal relationships or strength. To narrow it down, in this thesis I use the word cause to refer to physical, but not social, collective or psychological influence. Correspondingly, the term physical world refers to inanimate and animate matter, which is governed by regularities and physical forces (see e.g. Kistler, 2006; Crane & Mellor, 1989). This distinction is supported by developmental studies, highlighting that even very young children are capable to distinguish between causal influence of agents (humans, animals) and non-agents (billiard balls), the former corresponds to the free will, the latter refers to external causal forces (see Leising et al, 2008; Saxe, Tzelnic, & Carey, 2007; and Rottman, Ahn & Luhmann, 2011 for a review on the violation of causal determinism).
Throughout this thesis the word causation means relation between cause and effect (Kistler, 2006), while a cause refers to initiation that makes an effect happen. Originally, it is a philosophical debate as to whether causation is mind-independent, where relations are afforded by mental representations. One of the pioneers, David Hume (1739/1978), proposes three necessary conditions for this affordance: *contiguity and succession* in space and time; *precedence* between cause and effect (temporal asymmetry), and *regular repetitions* (the same cause must produce the same effects). In this account, a cause is about an object precedent and contiguous to another, and so united with it in the imagination that the idea of the one determines the mind to form the idea of the other (see also Hume 1711-1776/2007). This definition qualifies that a cause is actually a part of a larger entity, requiring a number of conditions involved when an effect is caused (Sloman, 2005).

Although being highly influential in developmental research (see e.g. Gopnik & Wellman, 2012; Gopnik et al., 2001 for learning causal mechanisms from regularities and Bayesian inference), the Humean account seems to neglect some forms of causality in the physical world. For instance, causation does not necessarily stipulate the involvement of two distinct AB type events, as it is studied largely in psychology literature (a hammer smashed a nut). Instead, in some instances causal relations can be intrinsic, embedded in temporally continuous processes, where the first cause is ambiguous, or a generating mechanism is more informative (the Earth travels around the Sun; in the context of electric charges, electrons flow through a continuous circuit). This neglected form of causality fundamentally differs from the Humean account due to its appreciation of the notion of space and time. For instance, guided by Salmon’s (1984, 1998) distinction between processes and events, causal events (one ball hitting another and making it move) are relatively localized in space and time, and the experience of contiguity marks interactions. One can think that, in principle, continuous processes can be conceptualized as a successive chain of contiguous events. However, processes are not perceived as such chains, but as continuous over time. In that sense, causation can be primarily a characteristic of continuous processes rather than a relation between AB type events. Because of this, in space-time diagrams events are shown by points, while processes are shown by lines, implying that causal relations are actually processes which generate
observable events (see also Dowe, 1999; and Reichenbach, 1958 for the distinction between genuine and pseudo processes –unreal sequences-).

Sowa (2000) argues that events are also governed by internal processes. The distinction is that processes can be described by starting and ending points in between while the change occurs. The characteristic of this change in the case of events is more discrete. A continuous process implies incremental changes from the start to the end. This suggests that when a continuous process is initiated (initiation), the effects continue to occur (continuation) until the ending point (cessation). A discrete process instead relies on the sequences of states as shown in Figure 1. In the example of the weather condition, the temperature varies by continuous gradations, but it is possible to take certain time periods and classify them as hot or cold. The distinction between discrete and continuous processes is persistent in mathematics (e.g. discrete versus continuous processes), in physics (e.g. four dimensional space-time continuum, electric field, magnetic field, quantum mechanical field), and also in computer science (e.g. discrete processes are defined by Pearl’s directed acyclic graphs –DAGs-), but in psychology the distinction or its meaning is poorly understood.

![Figure 1. The characteristics of processes (obtained from Sowa, 2000)](image)

There are other studies, following a non-Humean account and focusing on event cognition, most prominent in the Michottian tradition that examine whether causal relations can only be inferred from regularities as Hume suggested (e.g. see Leslie, 1984; Oakes and Cohen, 1990). These studies underline that in some circumstances the perception of causality in object interactions can be direct and immediate, without the assistance of prior experience, language, or causal learning (see Scholl & Tremoulet,
This strand of work suggests that there are other alternatives to understand causal connections, distinct from data intense approaches. However, in this strand of work neither invisible components nor causal mechanism is the concern. Though, this strand of work contributes to the literature in an important way by providing evidence for the distinction between perceptual and inferential processes involved in causal thinking, and thus it makes sense to bear in mind that causal reasoning is not a unitary process, but instead perception and inference require different cognitive resources (Fugelsang et al., 2005; Rolfs et al., 2013; Roser et al., 2005). However, in both strands of work (perceptual and inferential) the lack of methodological variety still restrict us, since the tasks that are employed to examine either perceptual or inferential causality typically rely on discrete AB type event approach. Moreover, in this framework spatial-temporal qualities have been commonly utilized in a discrete way, rather than continuous or unitary fashion. It is unknown whether causal relations in the physical world are conceivable in particular when spatial-temporal qualities are represented via continuous protocols rather than discrete.

The general aim in this thesis is to provide a developmental data set to further our understanding on whether humans can conceive cause-effect relations in the physical world. In particular, the objective is to examine what inferences individual children can make about causal processes from direct observation and comparisons as opposed to what they might have learnt or already understand about such processes, with a specific focus on (i) their capacity to do so in relation to continuous causal processes, which have been largely neglected in past research, and (ii) the ways in which this capacity is supported by how information is presented and by different cognitive capabilities –most notably, the ability to analyze spatial-temporal representations–.

Clearly, this thesis departs from the Humean paradigm, and for the first time examines the development of cognitive processes involved in reasoning about continuous causal processes. This departure presents challenges to the work here in two ways. Firstly, the literatures on both spatial-temporal cognition and the process view of causation are very limited, probably for the reason that the notions of ‘contiguity’ and ‘covariation information’ have received more interests in the than the notions of ‘continuity’ and
spatial-temporal analysis’. Secondly, describing or studying continuous processes in the ordinary terminology of cause and effect is highly difficult as Mackie (1980) argued. To overcome these difficulties the first chapter will provide a theoretical framework to define and discuss the key concepts, rationale, corresponding literature, and highlight the gap in the literature by discussing how reasoning about causal processes of natural phenomena is inadequately studied. The next three chapters will introduce how children’s ability to think and explain continuous processes develops and what kind of abilities are associated with understanding continuous causal processes.

1.2 The rationale of the thesis

1.2.1 Overview of the targets and hypotheses across the chapters

This thesis aims to provide the first systematic investigation on whether or not children can acquire causal knowledge from observation of natural phenomena relying on continuous processes without intervention. Two objectives serve for the general aim of the thesis, as elaborated below.

(1) The first objective seeks for an answer to the question of “Whether causal processes, in particular the mechanisms of natural phenomena (i.e. how a process works) are within the reach of children’s competences?” As outlined below, the developmental research in both scientific and causal reasoning strands has primarily focused on children’s ability to reason from evidence to the identification of causal variables, as well as their grasp of causal mechanism. Research has shown that children are capable of this kind of thinking from pre-school age. However, this literature has mostly involved reasoning about causal events in simple machine/toy like systems with distinct parts, or narrative approaches as discussed further. But understanding of causal processes in natural phenomena has never been studied. What we do not know is how reasoning about natural causal processes in such phenomena develops, or more specifically what we can learn from children’s reasoning about causal processes. This required a careful investigation, where children were tested individually on a series of causal tasks, which were chosen from natural phenomena relevant to physics (sinking), biology (absorption), and chemistry
(dissolving), using a novel approach in which they observed contrasting instances of each to promote attention to mechanism.

(2) The second objective aims to answer the question: “What kind of abilities are associated with children’s thinking of continuous causal processes?” This target is basically for understanding more about the development, the characteristics of thinking about causal processes, and its predictors. To achieve this, (i) we need to acknowledge what children already know about different aspects of the three scientific phenomena – sinking, absorption, dissolving– (prior knowledge), what happens when they observe and describe them (description from current observation), and what happens when they attempt to evaluate these causal processes explicitly (inference). Separation of prediction, description and explanation level responses helps to avoid potential pitfalls of working with verbal and nonverbal data and distinguish between ‘what’ and ‘why’ level responses. Note that the focus is on children’s reasoning not learning, and the methodology section below provided more details on these acknowledgements. We also need to (ii) understand what kind of cognitive abilities are linked to reasoning about causal processes when the tasks do not rely on laboratory-ruled causality or are not manipulated by a machine. Further tasks therefore assessed spatial, spatial-temporal, statistical, verbal (expressive vocabulary) and non-verbal (block design) ability to explore the key factors in children’s progress towards thinking about causal processes. The evidence from the event literature informs us that children show a kind of immature engagement with causal reasoning prior to school entry. Whether this ability develops with the same rate when causal processes are considered, or whether this ability develops into a mature form across the primary years is unknown. In particular, what competences for thinking about the natural causal processes develop and when? Answering these questions will aid research on how important distinguishing between ‘immature engagement with causal inference’ and ‘mechanism understanding’ is in scientific thinking.

The thesis consists of five chapters. The first introduces the rationale of the thesis, and the last discusses the overall findings obtained from the next three studies, each revolves around the above two questions. The second chapter investigates whether children
can infer causal mechanisms from observation of natural causal processes without intervention, and whether their capacity to do this is affected by how information is presented. The aim of this study is to establish a background to the natural world causal tasks, which will be used during the next two chapters. This study will test the hypothesis that children’s inference of causal relations is affected by information presentation.

The third chapter focuses on the developmental trajectories of children’s reasoning about continuous causal processes. It investigates the role of spatial, spatial-temporal, statistical, verbal, and nonverbal abilities. The aim of this study is to distinguish between prior knowledge, descriptive, and inferential level responses, and investigate which ability corresponds to these response levels. This study will test the hypothesis that spatial-temporal analysis is the most reliable predictor of causal processes across primary school years.

The fourth chapter aims to replicate the results of the third chapter, with a modified task battery. This chapter also differs regarding the implementation, methodology, and the sample characteristics. This study samples children from a wide socioeconomic (SES) background, where children from low SES school catchments are equally represented. Similar to the previous chapter, this study will test the same hypothesis and answer the question of whether spatial-temporal analysis is the key for conception of causal processes above/beyond statistical, spatial, verbal, and nonverbal abilities, and how these abilities evolve across the primary years when children’s school environments ranged widely.

1.2.2 Overview of the methodology

The participants were children aged 5–11 years with wide ethnic/linguistic variation, drawn from the different socioeconomic status (SES) range. The reason for working with this age group is threefold: first, in both scientific and causal reasoning literatures evidence consistently suggests that children start thinking about abstract concepts such as time from around the age of 4 and onwards (see e.g. Blakey et al., 2018; Hoerl & McCormack, 2018; Friedman, 1986, 1992; Lohse et al., 2015). Given that one of the
main concerns of this thesis is to understand the role of spatial-temporal analysis and compare it with other inferential tasks, such as covariation and probability judgments, the appropriate age to trial the tasks seems to be the primary level. Working with this age group can constitute a reference point for the researchers to modify the tasks for older and youngers participants at later stages.

Secondly, a large body of research sampled either preschoolers or adults probably for several reasons. For instance, the majority of the developmental changes have been seen during early years, and research typically aims to reveal the earliest detectible components of a particular ability to monitor the developmental phases more reliably. This monitoring is highly useful when the outcomes are compared with the adult data, or when the focus is on revealing the mature form of particular abilities. The causal reasoning literature is congruent to this picture, since the literature primarily involves early childhood and adult studies. However, for an unknown reason, primary age is quite neglected, leaving us with the limited knowledge on how reasoning abilities for causal processes evolve across mid and late childhood, or when actually required abilities mature into adult form.

Third, by sampling the primary age groups we can understand better what drives the development of reasoning about causal processes beyond language competences. The fact that language is one of the main progressing domains that can potentially affect task performance, in particular young children’s and preschoolers’ performances as discussed in the last section of this chapter (see section 1.5 explanation and causal reasoning with words). The nature of the causal tasks used in this thesis required children to conduct daily conversations. The literature assures us that although this ability continues to mature, it is typically well developed during the primary years. Therefore, daily language competences should not be a confounding factor, instead constrains must potentially be attributable to other abilities. This justifies why spatial, spatial-temporal, and statistical tasks were included in the batteries beyond verbal and nonverbal measures. Reducing the effect of basic language competences on task performances is highly crucial given that it is a mystery whether the development of reasoning about continuous causal processes is a
heterogeneous process by which various specific domain competences systematically develop, or it is dominated by basic language abilities alone.

Previously, there was not an appropriate methodology to investigate children’s understanding of natural causal processes. As briefly outlined below and detailed in the following chapters, the majority of psychology studies employed tasks where event A made event B happen in a distinct fashion (e.g. a cue ball is hitting a billiard ball, a plastic ball triggering the bell). The tasks used in previous studies required participants to employ spatial-temporal information from presented mechanisms with toys or mechanic models. Other studies were concerned with when and how causal relations are identified from co-variation, association to causation, or how people infer causal relations from contingency information. The tasks used in these kinds of studies typically focused on the scenarios of, for instance, causal dependency between cancer and smoking, the link between aspirin and headache, or temporal precedence between causes and effects as discussed broadly in the fourth chapter. On the other hand, the core idea of the process approach, as this thesis focused, is that causality needs to be observed or interpreted from the continuous interactions or quantity/quality transfer from cause(s) to effect(s). Exploring cognitive processes involved in this kind of inference is pivotal, given that most generative and preventive causes of the real-world phenomena unfold in continuous time. Giving some examples from various domains, for instance, protein folding is a process that cannot be grasped if transitions are separated based on spatial and temporal gaps. Similarly, physical growth does not occur in discrete fashion or in chunks. In line with these, evolutionary processes can be understood by acknowledging the flow of time. To address this core idea, in this thesis the three phenomena were presented in a way that the generative causal processes progressed continually in a reasonable testing length. For sinking, children observed two objects sinking in different rate in a long transparent jar filled with still water. For absorption, they observed two kinds of paper soaking up water in a certain period of time. For dissolving, children witnessed two kinds of salt disintegrating into water again in a reasonable time.

Each of these demonstrations consistently used contrasting instances (i.e. a stone and a berry/grape for sinking; a piece of tissue and blotting paper for absorption, table salt and
rock salt for dissolving) in order to make processes more salient for children as discussed in the second chapter. The focus was on children’s capacity to address variation in a single feature of each process, namely it’s speed, because speed is a characteristic feature of continuous processes. Processes differ in the speed with which they unfold – sinking is generally faster than absorption, which in turn is faster than solution – and this may affect children’s ability to observe and infer from observation. More importantly, there are differences in speed within each process, depending on the objects involved – a stone sinks faster than a berry, tissue paper soaks up water quicker than the blotting paper, or table salt dissolves faster than rock salt –. The causes of these differences relate to the underlying mechanisms. In the example of sinking, for instance, the stone sinks faster because the imbalance between liquid and object density is greater. Witnessing this variation in speed may alter the experience from a simple instance of a class of phenomena to one of related but quantitatively contrasting instances, leaving this difference to be explained. The use of minimally contrasting cases to highlight a problem dimension derives from perceptual psychology (cf. Chapter 2, see also Gibson & Gibson, 1955) and is frequent in science education (see Schwartz et al., 2011). Here, these contrasts should help children structure their observations, orienting them towards the target mechanism in an otherwise uniform process.

Investigating children’s ability to recognize of these processes required careful assessment protocols. Two approaches were employed to test the robustness of the outcomes as below:

(1) Initially, children’s responses were scored and analyzed based on a *three-stage structure*: prediction, description, and explanation (see Chapter 2 and 3, for further information). This protocol allowed children to inspect the materials and (i) predict what would they think to happen before they were presented with the mini experiments; (ii) watch carefully and describe what they actually saw; and (iii) explain why they thought things had happened the way they had seen. Children were encouraged to provide full answers at each point and at the end they were asked if they had more to say.
The replication study presented and explained in chapter four employed a five-stage design and extended the methodology analogous to the ‘scientific investigation process’. Children’s responses were scored and analyzed based on the five-stage approach, which consisted of systematic observation, prediction, testing, justification, and explanation. For each phenomenon, children first (i) were presented with the actual demonstration and observed the contrasting instances; then (ii) they were presented with three further instances and asked to predict the outcomes; (iii) they were also shown what happened to the three further instances to clarify whether their predictions were correct to enable them (iv) to justify their predictions, whether or not correct, and (v) explain why things had happened the way they witnessed. Full answers at each point were also encouraged.

The two approaches differed from each other in various ways: (i) the five-step design followed the actual scientific inquiry procedure, but without allowing children to intervene. Unlike the three-stage design, observation came before prediction, thus prediction did not rely on mere prior knowledge. This was useful to evaluate the role of prior knowledge before the observation, and application of their prior understanding to a novel instance. In the five-step design (ii) children were able to test their hypothesis, even justify their choices after testing, allowing them to evaluate/alter their views. Therefore (iii) these steps enabled children to get acquainted with the phenomena before the explanation. The same process was repeated across the implementation of the three phenomena. Finally, (iv) the five-step methodology allowed more room for non-verbal responses. Children observed various instances at two time points.

In both methods, children were encouraged to think about variants (i.e. water as the mediator, and objects), observables (i.e. shape, size, length, weight) and unobservable intervening factors (i.e. density, gravity, liquid density, compactness) to grasp the processes generating the outcomes and the way that they occur. The target was to understand how children move from thinking about variants and observables to thinking about processes with an awareness of unseen mechanism; one to the other stage of reasoning about causal processes. Although literature suggests that domain specific knowledge and early emergence of causal perception is crucial to mechanism understanding, though this does not explain our knowledge of the abstract nature of
causal phenomena. Moreover, research on invisible elements of causation has been primarily shaped by naïve theories (see e.g. Wellman & Gelman, 1992; Carey, 1985), evidence evaluation (e.g. Baldwin et al., 1993), causal learning (e.g. Gopnik et al., 2001, 2004), and Bayesian accounts of inferring abstract causal laws from minimal data (e.g. Schulz et al., 2008), but has neglected the role and the development of spatial-temporal analysis. Research also falls short in comparing the role of other candidate abilities in predicting causal processes/mechanism thinking as discussed further in the fifth chapter.

1.3 The relevance of scientific reasoning literature

Adopting Kuhn’s (2010) definition of scientific reasoning, it is about knowledge seeking, and does encompass any kind of purposeful thinking that enhances the seeker’s awareness. This definition proposes that scientific reasoning is something people ‘do’, not something people ‘have’. The latter is more about scientific understanding which is a product of the former. Bearing Holyoak and Morrison’s (2005) distinction, scientific reasoning involves the systematic transformations of mental representations, where various reasoning types (e.g. deductions, inductions) are involved when we think about scientific content, such as forces, species, or chemical processes (Dunbar & Fugelsang, 2005).

Scientific thinking, as a core subject in psychology, involves studies of cognitive processes, corresponding particularly to metacognitive modes of thinking, which can be used in all aspects of science. The metacognitive awareness enables reasoners to evaluate the epistemological status of the knowledge, with the support of evidence. According to Popper (1979) this is a process that requires continual effort to test theories against new evidence and falsify or clarify a proposal that allows for a solution. It is this metacognitive ability that enables humans to device the methods in order to find the ways to increase the harmony between theory, data, and method. This thesis does not speak to scientific reasoning in that kind, but instead provides data and evidence for children’s understanding of scientific topics.

Research suggests the involvement of various prerequisite skills in scientific reasoning, ranging from inquiry to discovery in the form of concept formation, or problem solving
Early research on scientific reasoning has encompassed two types of knowledge: domain-specific and domain-general (whether innate or learned). The former refers to knowledge about particular domains such as physics, biology, or chemistry, while the latter includes a complex set of abilities to support various science-related activities such as scientific discovery, search for a hypothesis via induction or abduction, or experimentation/hypothesis testing. The majority of developmental studies on scientific reasoning concerned one or the other of these two components.

The domain specific focus has its roots in Piaget’s pioneering work. In his 1936/1952 and also 1942/1974 books, he explored children’s understanding of physical and biological events. This developmental perspective motivated many researchers to investigate the distinction between domain-specific and domain-general knowledge to elaborate whether prior domain knowledge is likely to impose a strong bias when people reason about a real world phenomenon. Carey (1985) studied the development of biological concepts. Fay and Klahr (1996) examined young children’s understanding of determinacy and indeterminacy. Many researchers also focused on the knowledge independent tasks to reveal children’s ability to use domain-general reasoning competences. For instance, Siegler and Liebert (1975) investigated children’s
acquisition of experimental designs, specifically the planning and execution of factorial experiments. With knowledge-lean tasks the role of prior knowledge was minimized. Sodian, Zaitchik, and Carey (1991) showed that primary age children are able to determine which of two possible experiments can produce a reliable conclusion for the competing hypotheses. Kuhn et al., (1988); also Shaklee and Paszek (1985) used knowledge-lean tasks, they instructed participants to disregard their prior knowledge.

Research with adults focussed on people’s ‘knowledge’ about the natural world, such as pendulums, ball rolling, falling bodies, time, force, or temperature. McCloskey’s (1983) studies, for instance, showed that many well educated adults hold naïve beliefs about simple phenomena of motion. He found that these misconceptions were consistent with a pre-Newtonian, medieval theory of motion: the impetus theory. In his studies, McCloskey worked with undergraduate students and adults, asked them to predict the trajectories of moving objects (straight or curvilinear throws), and found that the majority of participants made frequent errors even if they take physics courses. He uncovered the bases of these errors and concluded that many errors about object motion are made, as participants hold a naïve theory of impetus. The naïve theory of impetus makes two assertions about motion: (1) the act of an object in motion passes to the objects via an internal force or ‘impetus’ that serves to maintain the motion. (2) The impetus gradually dissipates as a result of external influences (note that classical physics states that no force is required to keep an object at rest and no force is required to keep an object in motion). The difference his studies made is that he uncovered how people develop their theories of motion, what sorts of knowledge are acquired through experience with moving objects and what are the bases of their errors. His examples underline that naïve impetus theory is also very strongly held by children (see Kaiser, Proffitt, McCloskey, 1985) and is not easily changed during physics classes. These studies show that more accurate answers are likely among those who have systematic science lessons; some beliefs for object motion cannot be overcome via perceptual experience, but instead require formal instruction (see also diSessa, 1993; diSessa, Gillespie, Esterly, 2004, for further discussions).
McCloskey’s studies also show that explaining movements of physical objects often requires the combination of theories about forces and motion. His explanation is in accord with Piaget’s (1929) who proposed that children expect a ball rolling off the edge of a table to fall straight down. In Piagetian account, this misbelief arises from the lack of abstract thinking about the two vectors determining the ball’s motion. The misbelief occurs when children focus only on one force vector at a time (see also Viennot, 1979). However, White’s (1983) studies did not support this conclusion, as she found that even adults perform poorly on the similar force interaction tasks. Gilden and Proffitt (1989) also found that adults with naïve physics knowledge used one of the two alternative heuristics. They either considered the angle or velocity information for a similar task. Working with 160 elementary school children and also with adults, Pauen (1996) also found that most children and adults typically rely on one force only in their reasoning and prior knowledge about basic principles of forces does matter (see also Larkin, 1985). These follow up studies highlighted that either children or adults need formal training to understand force vector interactions.

Sampling much younger children (7-month-olds to 6-year-olds), Kim and Spelke (1999) studied the development of understanding of gravity and inertia on the basis of object motion. The participants needed to predict where a ball would land if rolled off a downward ramp. The experimenter showed three possible landing locations (straight-down, parabolic, and another path implying no gravity) and found that younger children up to 4 years of age consistently predicted a straight-down location, while 6-year-olds judged a parabolic path. The authors suggested that young children showed sensitivity to gravity, while elders showed sensitivity to both gravity and inertia. There are other studies showing infants’ use of physical and spatial information in the context of object permanence (i.e. Baillargeon, 1991; Baillargeon & Graber, 1987). Although these studies are not directly relevant, the findings support the idea that even infants show sensitivity to possible and impossible events and can reason about hidden object.

Wilkening’s studies shifted away from diagnosing the correctness of the responses, and allowed more space for children’s intuitive physical knowledge. His work pioneered studies which found interrelations and also dissociations between sensorimotor,
perceptual, and cognitive components of children’s physical knowledge in various domains, such as time, speed, force, and velocity (see Anderson & Wilkening, 1991; Krist, Fieberg, & Wilkening, 1993; Wilkening & Cacchione, 2011). These studies will be discussed broadly in the third and fourth chapters together with other studies (e.g. Wellman & Gelman, 1998; Bullock et al., 1982) focusing on the general issues of knowledge acquisition. The major contribution of this class of work is that the majority of them considered nonverbal thinking, which enabled comparisons across cultures, and elaborated the role of different forms of knowledge acquisition, such as implicit and explicit.

Other studies investigated the interactions between domain knowledge and the discovery processes by using laboratory simulations of real-world events. One of the pioneers, Klahr and Dunbar (1988) argued that science should be characterized as both product and process, and therefore the separation between knowledge –concept formation or knowledge acquisition– and strategy –experimentation, or evidence evaluation– in scientific reasoning is not realistic. They proposed a framework to show how these knowledge types can be integrated. In their general model of Scientific Discovery as Dual Search–(SDDS) the authors showed how scientific reasoning could be conceptualized as a search through two problem spaces: a hypothesis space, and an experiment space. The authors placed participants in a simulated scientific discovery process by first teaching them how to use an electronic device. They then asked participants to discover how a new function operates. The authors observed different patterns of search in two problem spaces and concluded that this dual search feeds into different strategies, which are highly crucial for abandoning or verifying hypotheses.

Schauble (1990) focused on children’s learning with self-regulated exploratory tasks. The author investigated fifth and sixth graders’ reasoning strategies over a long term program in which children worked over eight weeks (five hours in total) on a causal reasoning problem. The programme involved planning, performing, and learning about the relations between design features and speed of race cars in a computerized environment. The premise of the study was to understand how children revised their prior knowledge with the help of new evidence that they needed to create and interpret:
whether prior knowledge affected their reasoning (e.g. whether newly generated evidence is incorrectly interpreted when prior beliefs are inconsistent with it). In her investigation Schauble found that children tended to interpret the same evidence differently, depending on whether the evidence confirmed their prior knowledge or not. Very few children could coordinate their prior beliefs with new evidence. In a follow up Schauble et al. (1991) also argued that the discovery process is driven by developing links between knowledge and the process of experimentation.

As Zimmerman (2000) discussed, various studies repeated these results and examined the effect of knowledge on reasoning. They either minimized prior knowledge with knowledge-mean tasks or instructed participants to disregard their prior knowledge. However, these added little to our understanding of the development of scientific reasoning. Current approaches shifted away from that focus and concerned more with how research can help children to become better in science. For instance, Chen and Klahr (1999) investigated the impact of direct and indirect instruction of the control-of-variables strategy. Kuhn (1989), Kuhn, Amsel, O’Loughlin (1988), Kuhn, Cheney, Weinstock, (2000) investigated when and how children develop specific scientific reasoning strategies, such as knowledge seeking, acknowledging one’s existing knowledge, experimental strategies, hypothesis testing, correcting incorrect or incomplete information, developing the theory-of-mind, and coordinating theories with evidence in an intentional and explicit way. These studies highlighted that children show difficulties in coordinating two phases, namely evidence collection and construction of appropriate theories.

According to Kuhn (2007), various scientific skills develop late. For instance, in an extended intervention, children were asked to focus on a problem on a regular basis over a period of three months. Two strategies were promoted: control-of-variables, and multivariable-prediction. To promote children’s ability of control-of-variables, children were asked to identify five causal and non-causal variables influencing the outcome. Later, children were asked to predict outcomes for novel variable situations, and justify their predictions (multivariable-prediction). The target was to assess whether successful children in the first activity were also successful in the second. The majority of children
isolated the causal and non-causal variables operating in a multivariate system well in the first activity. When they were asked to predict the outcomes of various constellations, they did not consider the effects of all causal variables they had identified. Moreover, they did not consider whether their predictions were consistent with the previous identifications. This study suggested that children’s mastery of the control-of-variables was the key to success in their scientific reasoning. However, most aspects of scientific thinking seemed to undergo developmental processes, such as experimentation, evidence collection, evaluation, argumentation, and epistemological understanding of the nature of science (see also Dean & Kuhn, 2007; Kuhn, et al., 2008; Zimmerman, 2000, for a review). This literature confirms that maturation in these capacities is attributable to development of internal metacognitive abilities.

In the end, two main conclusions can be drawn from scientific reasoning literature. First, as Klahr and Dunbar (1988; see also O’Brien, Costa, & Overton, 1986, Zimmerman, 2000) argued, it is plausible to think that the separation between the two classes of investigation –strategies versus knowledge– seems artificial. Studies following any of these strands inform us that knowledge and investigation strategies mutually influence children’s competences of interpretation, hypothesis generation, experimentation, and result generation. This conclusion is informative for the first target of this thesis (whether causal processes are within the reach of children’s competences) in the sense that eliminating prior knowledge from reasoning is compelling, since knowledge grows and specialized within various domains from very early on. Children acquire knowledge from their surroundings, and exposed to various physical, biological, chemical incidents promoting that growth, and possibly affecting their thinking strategies. The second conclusion is that the above studies have mostly focused on general principles invoked in reasoning about scientific events. By contrast, there are other studies differentiating scientific conceptualizations further under physical, chemical, and biological domains. The next section will give an account of the progress the fields have seen recently in these domains. To narrow down, this review will focus on three phenomena -sinking, absorption, solution-, as studied in this thesis. This review will show that although these studies did not concerned with reasoning about causal processes, they shed light onto the notion of children’s understanding of scientific topics and thus elucidated what we already know about their thinking about those phenomena.
1.3.1.1 Previous studies on sinking, absorption, and solution

1.3.1.1.1 Studies on sinking (physics)

In the domain of sinking Piaget and Inhelder’s (1948/1974) work pioneered the idea that density is a complex concept, and from the age of 7, children show a cognitive shift in their mental operations where they become aware of categories and classifications (see Driver, 1985; Smedslund, 1961, for studies investigating children’s concepts of weight). Smith, Carey, Wiser (1985) criticised Piaget and Inhelder’s complex tasks given to young children and conducted a series of experiments to elaborate the developmental courses of some relevant concepts, such as size, weight, and density. Children were asked to make a clay ball that would weigh the same as a wax ball: size was intruded upon a weight judgment, and success on this task required the conceptualisation of density. The authors asked children which out of a pair of objects was made of a heavier stuff. The pairs contained small but denser materials and a large relatively less dense material (e.g. cubes made of aluminium, plexiglass, wood, steel, brass etc. in different size). Young children’s failure showed that they systematically mixed up weight with density. Children aged 5- to 7-years were able to modify the concepts but the majority of them indicated that heavier objects were made of heavier stuff. In this study, critical pairs were shown by pitting weight against density, thus children’s knowledge about the density was assessed. The authors found that right judgments appeared after the age of 8 and onwards.

Kohn and Landau (1987) extended this target and explored whether children’s knowledge of density was influenced by the task design, and tested this idea by also focusing on floating. Children, 3-to 5-year-olds were shown a series of 8 objects (dense: metal and clay; not dense: plastic and wood; object type: ball, spoon). The children were asked to judge whether these objects would sink or float, and explain why. The authors found that 5-year-olds performed almost perfectly, much better than the 3-year-olds, judging the majority of the trials correctly. Children’s errors were not related to object type, size, or weight alone, but instead the author found that the nature of the substances was another factor that affected children’s judgments. For instance, the majority of the
errors were made on wood and metal objects: most children judged that the heaviest/largest wooden objects would sink. In the same study, children’s performances were compared with adults. Regarding the number of correct predictions, the adults performed better than the children. However, both adults and children made the most errors on the same object – wooden pieces –, highlighting that some objects may particularly confuse children in relation to mass, volume, and density (see also Smith, Maclin, Grosslight, Davis, 1997).

There seems to be two possible ways to explain children’s difficulties with density. The first follows Piaget and Inhelder’s conclusion proposing that children have difficulties in integrating dimensions until the formal operational stage. The second is that children have difficulties in comprehending the questions about density and relate other properties, such as weight (see Kohn, 1993, for a review). To test whether children’s performances in previous studies were dependent on the experimenter’s explanation about the density (e.g. ‘which one is made of heavy stuff’ or ‘heavy for size’), which might be confusing in particular for the young children, Kohn (1993) designed another experiment, and sampled a wide age range, children from 2- to 11-year-olds. Kohn made certain that the children, particularly the preschoolers, understood the question well and they knew they were not being asked about weight. She used the concept of buoyancy as a crucial consequence of density, using a much simpler version in which children needed to judge if sets of 18 blocks (made of e.g. wood, aluminum, balsa wood blocks) float without explaining why. The author found that the errors made by children and adults with particular objects were similar to errors reported in Smith et al.’s (1985) and also Kohn and Landau’s (1987) designs, where weight and volume intrude systematically. However, children as young as 4 years showed some conception of density that allowed them to predict the buoyancy of the objects. The data suggested a continuous developmental trend: although young children might lack a full comprehension of density, they develop some conception around the age of 4 or 5 about it. Though, adult’s performance on the same tasks were not perfect neither. Kohn’s (1993) study showed that children demonstrate a differentiated density concept around the age of 5. Note that Smith et al.’s (1985) study concluded that children come to differentiate weight and density between the ages of 8 and 10. Kohn’s study proposed that young children could understand how the notions of mass and volume are related to object’s sinking rates.
This point was explained by the authors with the possibly that children would be highly sensitive to physical regularities.

In these experiments, the other relevant factor was the size of objects. Most children, in particular the young ones, seem to believe that larger objects sink faster than smaller ones. Kloos, Fisher, Van Orden (2010) compared children’s and adults’ understanding of density to elaborate on whether naïve minds are riddled with misconceptions due to the nature of constraints available in different experiment protocols (e.g. actors, tasks, saliency), or their performance differs as a function of their knowledge. In this study, the first experiment replicated Kohn’s (1993) design, where participants were asked to predict whether cubes would sink or float. The difference was that they changed the experimental protocol to test the saliency of object density. A group of children were shown objects in pairs (salient category), while the other group were shown single objects in each trial. They found that adults’ and young children’s judgments were affected by the trial protocol. These groups performed at chance for single presentation trials as compared to the pair presentation trials. In the second experiment they investigated whether children’s performance on single presentation trials suffered from boredom, or other confounded factors. Children were presented with less trials (30 in total) in which they needed to sort cubes into sinkers versus floaters. The authors found that many 4- to 5-year-olds were successful in this protocol, replicating Kohn’s (1993) finding. However, the sample of this second experiment was inadequate and included only 9 girls, and 11 boys, all pre-schoolers. In the third experiment, the authors presented children with pairs of cubes only, with one type of pair, where mass correlated with density -the heavier object in a pair was the denser one-. The procedure was identical to the first experiment (Kohn’s protocol), and the authors found that the children, even younger ones, easily distinguished between sinkers and floaters in a pair condition when density and mass correlated positively (heavier objects are the denser). This highlights that low/high performance in the pair presentation trials may be due to consequence of saliency in task manipulations.

Working with college students Mullet and Gervais (1990) asked 12-, 13- and 15-year-olds to estimate the mass and weight of objects. They found that given only abstract descriptions of the objects the majority of students possess two distinct concepts about
mass and weight, but they are unable to distinguish between those two concepts. Children used incorrect rules to combine information about volume and density, implying that weight can also be a difficult concept when it is pitted against mass. Carey (1991) also discussed that weight and density might not be easily differentiable concepts for either children or adults due to the fact that they are interdependent concepts.

Penner and Klahr’s (1996) study differed from the above studies in the sense that they treated density as one of many factors that determine how quickly an object would sink. The authors aimed to understand whether domain-specific knowledge would interact with domain-general scientific discovery skills (i.e. with experimental strategies). In the study, the domain-specific knowledge was explored by sinking. Children, 10- to 14-year-olds, were asked to explain why objects sink and which attributes contribute to sinking rate in water. They were then shown six objects (e.g. steel washers, a rubber mouse etc.) to talk about. Then, the participants were presented with 15 possible pairs of objects one at a time, and they were asked to predict which objects would sink faster in water. The participants explored their hypotheses by dropping the objects (singly or in pairs) in water, and then they were asked to conclude what factors they believed were important to the sinking rate. The authors focused in particular on the role of experimentation and hypothesis testing in inference/knowledge revision. To assure this, comparison of the number and type of attributes mentioned by each participants before and after experimentation. The major finding of this study was that although the majority of participants focused on the role of weight in their predictions, after the experimental phase most of them declared that weight alone was not a sufficient factor for determining sinking rate. However, their data is not conclusive as to whether change in participants’ statements was due to experimental phase, or whether this phase helped them to revise their statements or remind them of other attributes.

Hardy et al. (2006) investigated how 9-year-olds could revise their misconceptions about sinking and floating as a consequence of instructional support. The authors found that teaching these concepts fostered a long-term reduction in misconceptions and improved children’s scientific explanations. There are other studies exploring, for instance, the effect of peer interaction in the context of sinking and floating (Howe, Tolmie, Rodgers,
1990), or effect of instruction in young children’s understanding of sinking and floating (Rappolt-Schlichtmann et al., 2007). However, the majority of current studies mostly replicated or followed the routes of the above mentioned studies (see e.g. Meindertsm et al., 2014; Schneider & Hardy, 2013). The essence of these studies is that the notion of density emerges from an initial undifferentiated weight-density concept, and therefore in the context of sinking, children must understand how mass and volume are related to an object’s sinking speed (see Sodian, Zaitchik, Carey, 1991; Kloos & Van Orden, 2005; Hewson, 1986; Kloss, 2007; Smith et al., 1985, for a review). Conceptualizing density seems to be relevant to the tasks of sorting objects regarding the material kind, salience manipulations, child’s knowledge, experience on floating and sinking, research question, and also understanding of ratio and proportionality.

In this thesis the focus is not on density concept, or sinking/floating alone, but instead on the determinants of sinking as a process. The physical process behind sinking is quite complex. When an object sinks, its acceleration is determined by the factors (weight, volume, size, density of the objects) highly relevant to frictional force, and also other variants such as gravity and the countervailing forces of buoyancy. This implies that understanding of sinking requires multiple factors and variants to be considered in an explicit way. On the other hand, the literature highlights that children may not be good at considering more than one kind of force/variant in their observation and the saliency is an important factor affecting their performance. To address these points and make the process more salient, the contrast pairs were used across all studies (e.g. a stone and a berry/grape). Children at primary level are not expected to provide a clear picture about this complex process, but the postulation here is that the determinants of sinking as a process are not completely knowledge-based. Instead, we can expect children to see the objects behaving differently in the same environment due to effect of the variants and intervening factors, such as gravity and liquid density in a more holistic fashion. This design can be seen as close to Penner and Klahr’s (1996) approach, but differ significantly from that study in the sense that the aim here is not to investigate children’s experimental strategies or domain knowledge interactions. Sinking is seen as a process representing observable causal relations in continuous processes that encompass unseen latent variables.
Literature on children’s knowledge/reasoning about absorption seems to be under investigated, leaving us with a lack of understanding as to whether the concept of absorption is a complex process that requires specific relations to be incorporated. Initially children’s understanding of biological phenomena was studied with regard to conceptual development (Wellman & Gelman, 1998). Carey’s (1985) study reviewed Piaget’s notion of animism and discussed the notion that although young children’s reasoning about biological phenomena is animistic, it is not due to their domain general intellectual maturity, but about lack of domain-specific knowledge. Children younger than 5 years old possess a theory-like knowledge system, which contains teleological and causal devices enabling them to predict and explain biological phenomena, which is called naïve biology. Children’s knowledge system enables them to distinguish between mind-body, and living-nonliving things (Inagaki & Hatano, 2002; see also Ericson, Keil, Lockhart, 2010, for further evidence on the distinct operations between biological and psychological processes).

Later, various studies showed a diverse interest, rather than coordinated. The focus in this domain varied from the categorization of students’ errors on real time cognitive tasks (Fisher & Lipson, 1986; Norman, 1981) to conceptual frameworks (Wandersee, Mintzes, Novak, 1994; Carey, 1987), or children’s misconceptions in biological events (Carey, 1985; Dreyfus and Jungwirth, 1989). Preschoolers’ views about biological concepts of animals received more interest. Carey (1987), for instance, discussed how children could construct a new intuitive framework theory of biology, with animals and plants consolidated into a core category of living things. She saw that across development, the concepts of person and animal changed in fundamental ways. Children showed change in their categorizations of the concepts based on underlining factors. For instance, a person and an animal can be distinct from other inanimate things based on the factors whether or not they are alive or dead. Simons and Keil (1995) examined along what dimensions young children expected animals to differ from artifacts, whether they had distinct explanations for biological (e.g. animals) and physical (e.g. machines) causal agents. The authors showed that young children’s expectations about the insides of animals and machines differed. By the age of 8, children showed sufficient concrete knowledge to
consistently distinguish between the insides of natural kinds (rocks or animals) and the insides of artifacts (machines or blocks). Miller and Bartsch (1997) compared first and third graders’ explanations of biological phenomena with adults’ and found that children’s explanations are not particularly more vitalistic than adults.

Regarding children’s understanding of body-mind distinction, Kalish (1997) found that preschool children distinguish between mental (emotional) and physical (illness) reactions to contamination: while eating contaminated food makes one bodily sick, knowing this induces mental reactions. Schult and Wellman (1997) found that preschoolers could differentiate and coordinate psychological (beliefs and desires), physical (forces), and biological (reflexes) processes of human movements. Gottfried and Gelman (2005’ see also Gelman & Kremer, 1991) investigated the developmental patterns of understanding biological causal agents. The authors found that young children would recognize the need for an inherent mechanism for behaviors of living things, but lacked specificity as to what that mechanism was. Increased conceptual knowledge aided children’s explanations. Even more, young children appeared to have impressive knowledge about the categories of plants, animals, and machines, enabling them to talk about things that are inside each kind of objects, such as gears and batteries that are inside machines, hearths and bones that are inside animals. These studies highlighted the importance of domain-specificity in children’s explanatory thinking, which is a departure from the Piagetian view that assumed young children to be incapable of offering a plausible explanation in many domains, including topics relevant to biology.

Legare, Wellman, and Gelman’s (2009) study provided further evidence for preschoolers’ readily available language for most biological facts. In this study, children needed to explain a biological phenomenon – contamination – and they did so, as the authors found that preschoolers had readily available contamination based explanations, even when an unseen mechanism (germ) needed to be considered (see also Siegal & Peterson, 1999; Keil et al., 1999). Springer and Keil (1991) also found that preschoolers are highly sensitive to plausible and implausible mechanisms for biological events. For instance, sun and rain can be causal agents in color acquisition of a flower, but a little man with a paintbrush opening the seed cannot be the reason.
Wellman, Hickling, and Schult (1997) proposed that young children develop three distinctive explanatory reasoning systems: psychological, biological, and physical. Explanation of human behaviour involves construing human action in terms of actors’ internal mental states; an early understanding of physical and mechanical correspondences; and an early understanding of daily biological states, such as birth, growth, or illness. The distinct reasoning systems are shaped through children’s conversational interactions. Hatano & Inagaki (1994) suggest that young children possess at least three components that constitute their thinking about biological phenomena: knowledge about the living-nonliving things and also mind-body distinction that enable them to specify objects; mode of inference which enables children to produce consistent predictions about biological kinds; and last one is non-intentional causal explanatory framework that young children use to organize the body of knowledge they have to determine their theories.

Adams and Griffard (2001) analyzed college students’ content knowledge on physical and biological phenomena. The authors proposed a matrix of attributes that characterizes various conceptions in the two domains. A comparison of students’ conceptions for biology and physics showed that: interconnectedness between the concepts played a major role in students’ physics and biology content knowledge, though in distinct ways. For instance, focusing on interconnected topics in physics such as force, energy, and motion would help students to change their misconceptions. Another finding was that the biology conceptions tend to be more language sensitive than in physics. This result is in consistent with Legare et al.’s (2009) and Wellman et al.’s (1997) arguments. As Adam and Griffard (2001) argued, one possible explanation would be that humans are living things and interact with other living things around their surroundings. The social aspects of human language play a major role, but the language seems to include various concepts particularly relevant to biological facts and events from very early on.

In the end, our lack of knowledge remains, as none of these studies compared children’s reasoning about biological, physical, and chemical phenomena. Moreover, in the domain of biology, we know very little about children’s thinking about absorption. The topics of
osmosis, capillary action, and internal transport in plants/humans can be seen as contextually the most relevant. If we look at the literature on children’s understanding of these topics, the closest studies can be seen as follows. By using concept maps and structured interviews, Arnaudin and Mintzes (1985) investigated twenty-five fourth graders’ knowledge of the human circulatory system, such as cardiovascular concepts, structure and function of blood/heart, closed circulation etc. In many instances, the fourth graders did not provide acceptable responses, particularly when conceptual maps were not in use. Odom and Barrow (1995) explored college students’ understanding of osmosis and diffusion, and explored the impact of classroom instruction. Wang (2004) developed a diagnostic instrument to investigate college students’ naïve conceptions, inaccurate or incomplete understanding about internal transport in plants and human circulatory systems. In his survey, he found that many Taiwanese students hold alternative conceptions about the internal transport in plants and humans. For instance, most students strongly believe that the energy of photosynthesis pulls water upward in the plant; nutrients made in leaves are turned into oxygen; capillary action occurs via plant roots, which absorb water through vessels. Driver et al. (2001) also argue that students have incomplete understanding about photosynthesis, overlapping with Wang’s findings in the sense that children have naïve understanding about the energy transfer in plant metabolism.

To sum up, most developmentalists seem to agree that young children possess theories in the domain of biology, where innate and cognitive competences are combined to think about e.g. the behaviour of animals, plants, bodily processes, or inanimate things. We do not know in particular when and how children get familiar with and reason about absorption or relevant phenomenon like capillary action, which is highly relevant to many biological phenomena ranging from blood circulation to vitality of plants, animals, humans. Note that the causal processes of absorption serves for many life functions, where adhesion of liquid, its molecules, and the nature of the vessel walls together with the surface tension cause an upward force. Many objects can also be utilized in similar processes, such as paper towels, sponges, filter papers, or the tips of fountain pens.
1.3.1.3 Studies on dissolving/salt solution (chemistry)

This domain also has its roots in Piaget and Inhelder’s (1974) pioneering work on the conservation of substance, weight, and volume during solution, where children were asked which properties would remain invariant when sugar dissolved into water. Piaget and Inhelder reported that young children up to 7 years were pre-atomist. For instance, they use the word particle in the context of grain rather than particles in an atomic sense, with very limited conception of the nature of dissolving. Since then, earlier studies investigated children’s conceptions of dissolving, but the focus was mostly on the idea of conservation of matter. For instance, Osborne and Cosgrove, (1983) interviewed 8- to 17-year-olds to investigate their conceptions about evaporation, boiling, and melting of ice. Moore, Dixon, Haines (1991) focused on the development of proportional thinking using a temperature mixture task. Working with 2nd, 5th, and 8th graders they asked children to predict the outcome of combining two containers of water at different temperatures, and found age related changes in children’s intuitive and computational knowledge of proportional reasoning.

Strauss and Stavy (1982) asked children about the sweetness of a mixture in two glasses of water, with the same or different amount of sugar dissolved in them. In this design, children younger than 5 year old provided correct answers by relying on their intuitive understanding (they seemed to take into account the quantity of sugar and water), while children between 6 and 10 years of age mostly tried to apply quantitative rules and ratios which caused their performance to drop (mixture would become sweeter than the two initial components). Children older than 10 seemed to correctly take into account the quantities in both components and how ratios were manipulated even when the amount of sugar and water varied. The authors reported an U-curve in children’s responses across development. Stavy et al. (1982) attempted to determine whether the U-shaped knowledge growth was specific to sugar-water solutions or a more general tendency that can be found in other domains. He reported that the U-curve is seen for the topic of temperature as well. The drop in the percentage of correct answers was interpreted as part of children’s judgments of quantities: when their quantity judgments were ordinal they needed to compare and order the relative quantities.
Driver’s (1985) study also argued that different factors might affect children’s uncertain beliefs about conservation and dissolving. If children are encouraged enough to think beyond perceptual appearances, this may improve their thinking about the sugar as being still present in the solution. Longden, Black, Solomon (1991) attempted to analyze 11- to 14-year-olds’ difficulties in recognizing dissolving even if the examples are relevant to their everyday experiences. Children were asked to draw and write about dissolving in two different ways. In the first, they needed to draw what it looked like when the crystal was dissolved. In the second, they needed to draw what the atoms looked like when the crystal was dissolved. First required everyday representations, while the second was described as dissolving in particle terms. The authors found that the number of children holding a correct view of dissolving at the everyday level was less than those answering the particle interpretation of dissolving correctly. The authors interpreted this result as a consequence of school teaching, and discussed that scientific concepts are embedded in everyday ideas as was shown in the example of dissolving. Therefore, everyday examples should not be excluded from science teaching.

From the students’ comprehension point of view, research indicates that some students are not able to distinguish between physical and chemical changes. For instance, Meheut, Saltiel, Tiberghien (1985) reported that some students do not consider chemical reactions as complete transformations of the matter, but instead they focus on the change in appearance. As a consequence, chemical processes can be misinterpreted as mixing the initial substances and their properties, but not change in the constituent. Johnson (2010) explored 11-to-14-year-olds’ conceptual understanding of substances and chemical change (melting and boiling). Although he considered children’s previous interactions with melting and boiling, he found that children did not naturally have the concept of substance identity, enabling them to recognize chemical changes. Using an interview technique BouJaoude (1991) found two reasons underlying adolescents’ lack of understanding of the burning process. He reported that students relied mostly on memorized information to explain an observed phenomenon, and held fragmented, inconsistent, and task-specific understanding. Ahtee and Varjola (1998) investigated 13- to 20-year olds’ understanding of the meaning of chemical reaction. He found that almost 20% of the 13 and 14 year olds and 17-18 year olds thought that dissolving and change of state were chemical reactions. Only one-sixth of university students could explain
what chemical reaction had actually occurred (see also Andersson’s 1986 categorization of students’ explanations about chemical change).

Other strands of research considered the characteristics of chemical phenomena and underlined that chemical events involve invisible properties and processes, and therefore explaining such processes may be particularly difficult for children. Flawell (1986), for instance, found that most preschoolers show no understanding of the appearance-reality distinction. Around the age of 6 or 7 they begin to make correct appearance-reality distinctions and start distinguishing between ‘looks like’ and ‘really and truly’. Rosen and Rozin (1993) examined preschoolers’ sensitivity to invisible particles in dissolving. The authors tested Piaget and Inhelder’s findings with a more age sensitive procedure, with the assumption that invisible processes may influence children’s judgments about dissolving of solutes (sugar, citric acid) into a solution (water). The authors hypothesized that a mature conception of the dissolving process should imply appreciation of the appearance-disappearance of matter when for instance sugar gives the water a property of sweetness. Although it is not possible to visibly trace the dissolving process itself -as the water would be visually identical to pure water-, there is a sensible trace of taste. Children were examined via six conditions in which sugar (first condition), citric acid (second), polycose (third), and laundry soap power (fourth) were stirred into a glasses of water (in the fifth condition the citric acid was placed outside the glass, not dissolved in the water. In the sixth condition randomly assigned subjects were presented with some of the previous demonstrations). The authors reported rather early maturation in the notion of appearance-disappearance of invisible particles. Contrary to Piaget and Inhelder’s (1974) interpretation, the authors reported that most 4-year-olds showed sophisticated realization of invisible particles, while much of the development in this kind of awareness occurs between 3 and 5 years. In particular, the citric acid and soap conditions indicated that although children used multiple sensory strategies to detect visually identical solutions, neither pleasantness, nor smell/taste significantly affected children conception about the nature of dissolving. However, many 4- 5-year-olds acknowledged the presence of the entity even in the absence of visual residue.
There are other studies, for instance, investigating the quantitative aspects of solution such as solubility (Gennaro, 1981); how external factors such as stirring and change in temperature affect the dissolution process (Blanco & Prieto, 1997); or how to help students to think beyond appearances (Kind, 2004). However, their scope is not as relevant to the focus of the present thesis. The emphasis here is not on children’s knowledge about salt solution/dissolving, but on their ability to observe and reason about dissolving, and certain characteristics of this process. A solution presents a specific type of mixture when a substance is dissolved into another. Neither solute nor solvent molecules can be observed by the naked eye, but actually the solute breaks up from a larger quantity into smaller individual molecules. In this thesis the nature of the solvent (water) was kept constant while solute (salt or sugar types) varied. Saturation was not a concern, as each time a tiny piece from each salt type was dissolved in fresh water. The demonstration process followed the same protocol as in sinking and absorption to allow comparisons on whether children’s development in solution is roughly similar with regard to inferring these phenomena. One can suggest that chemical phenomena may be more challenging for children. However, our lack of understanding remains, as no study has yet compared children’s explanations of different phenomena.

1.3.2 Contribution to the scientific reasoning literature

We have witnessed that various cognitive phenomena have been studied within the scientific reasoning framework. In particular, most previous research was concerned with when concepts like density would emerge and develop into a mature form. However, how such emerging knowledge affects, or is affected by, other competences is rarely investigated. The methodology of Penner and Klahr’s (1996) study is the most similar one with regard to one of the tasks employed in this thesis, namely sinking. The authors asked participants to drop items in water and explain the determinants of objects sinking rate. However, in their study Penner and Klahr focused on how children’s experiments in the micro domain of sinking integrated with children’s experimental strategies; whether children’s experimental strategies improved their explanations. The processes by which children come to understand from demonstrations rather than experimentation was not the focus. The methodology employed in this thesis is concerned in particular with whether children come to understand underlying causal processes of sinking, absorption, and solution without experimentation. One can suggest that Penner and Klahr’s study...
elaborates on whether children improve their understanding about a phenomenon by intervention/experimentation, while this study explores whether they can reason about underlying processes through observation without intervention.

Secondly, the concern throughout the thesis is not children’s prior misconceptions, as largely studied in scientific reasoning literature. For instance, in the literature the tasks used to test children’s understanding of density often involved scenarios in which variations in density were pitted against variations in mass or volume. For instance, children were presented with pairs of objects of which the heaviest one had the lowest density, or vice versa (see Kloos et al., 2010, for the classifications of these tasks; and also Hewson & Hewson, 1983; Kloos & Van Orden, 2005; Smith, Carey, Wiser, 1985, for the further examples). The results show that in particular young children are likely to perform poorly in this kind of task design. For instance in Smith et al.’s (1985) sample the majority of 3, 4, even 5-year-olds incorrectly concluded that the less dense object was made of heavy stuff. When an aluminium block was heavier than a steel block, more than two thirds of 8- to 9-year-olds categorised the density of the blocks incorrectly. To avoid the constrains arising from the task protocols, such as children’s possible misconceptions about mass and size to infer about density, in this thesis identical-looking objects were included. For instance, a berry or a grape is used with the same shape, size, and colour of a stone.

Another point is that although various phenomena were studied within the scientific thinking framework, we are yet to answer the question whether the most studied domain of sinking is easier for children as compared to biological or chemical processes. In other words, whether children use similar strategies to reason about different phenomena, even if the phenomena are presented in a similar way, or their reasoning about the determinants of the phenomena differs with the effect of other factors, either obtained from everyday experience, or through cognitive capacities. One can argue that children may witness more sinking related activities by observing objects in bathtubs, by swimming, or via drinks. Others can suggest that absorption is the most relevant to everyday life via cleaning activities, drying hands with fabric etc. Alternatively, others can argue that children must be more familiar with dissolving due to use of daily
substances like salt and sugar. Although the underlying physical, biological, and chemical processes of these phenomena are quite complex, previous work shows that children are able to talk about these kinds of phenomena even in a sophisticated way (e.g. Gelman & Wellman, 1991; Sobel & Legare, 2014). However, we have no data to further our understanding as to whether children’s specific domain knowledge consistently affects their reasoning. Or indeed, whether children in particular find a specific domain more challenging to grasp. This is highly crucial for school science, teaching and learning processes. The thesis aims to provide the first comparison on this point.

Another unique target is to explore to what extent children would think about abstract variables involved in causal processes by just watching objects themselves behaving differently in a shared environment. Obviously, neither the notion of density, absorption, nor solution is directly observable properties of the objects themselves; instead these are hidden features that need to be understood via interactions. In the contexts of sinking, for instance, density needs to be inferred from the relation of the mass per unit of volume, and therefore mass representing heaviness and volume representing size are more obvious features as compared to density. Such dimensional factor seems to make its grasp likely to appear late in the development. Therefore, Piaget’s (1964), and Piaget and Inhelder’s (1942/1974) stage approach may be factual at primary level, or confounded by the individual differences shaped by cognitive and environmental factors. This thesis will provide an answer to the question of what drives children’s intellectual maturity in the three domains. Is it mere cognitive capacities, as indexed by the IQ measures, or are cognitive capacities bound by other environmental factors when it comes to differentiating between, for instance, density and heaviness; spreading and dissolving.

1.4 The relevance of causal reasoning literature

At their roots, the theoretical goals of scientific and causal reasoning are similar: both aim to explain the phenomena, and enhance the explanatory power of the theories and models to control unwanted outcomes. However, the literature on causal reasoning and the literature on scientific thinking have almost progressed independently in a way that the two literatures differ with regard to the focus and the methodological tendencies. As
Kuhn and Dean (2004) underlined, causal reasoning literature largely stands to benefit from the identification of how people take advantage of regularities to induce causal relations or construct a mental model to improve their understanding about their surroundings. Koslowski and Masnick (2010) also highlight this puzzling disconnect in the psychology literature by stating that most research on causal reasoning defined causation in terms of Humean indices, such as temporal priority, contiguity, and covariation. Initially the majority of research followed the theoretical ideas of Hume, Mill, and more recently Kelley, and Pearl’s pioneering works. However, in Woodward’s (2016) words, the relation between causality and science is epistemological.

Both scientific and causal reasoning literatures encompass a broad set of cognitive phenomena from various domains, with an established interest in explicit/implicit forms of thinking. However, within the scientific reasoning strand there is a substantial interest in both developmental and adult literature. On the other hand, causal reasoning literature largely samples early childhood and adulthood age ranges. The reason why young children’s causal understanding received more attention than the primary or secondary age levels might be due to the link with corresponding discussions on the origins of causal knowledge, in particular engagement with the domain specific and domain general knowledge, as various studies show the role of conceptual development in causal thinking, such as McClelland and Thompson’s (2007); Oakes and Cohen’s (1990); Schulz and Gopnik’s (2004); Spelke’s (1990); Spelke et al.’s (1992); Spelke and Kinzler’s (2007) studies.

This thesis proposes that the ostensible disconnect between the scientific and causal reasoning literatures is not foundational. Although there is enough room for noncausal forms of thinking in both literatures, causal reasoning plays a central role in many areas of science, ranging from the social, behavioral, biological, statistical to artificial intelligence, with a general aim of imposing a valid explanation. The apparent disconnect disappears, for instance, when we consider the earliest developmental studies. For instance, Piaget’s (1930, 1970/1972) pioneering work distinguished between the three developmental periods of causal reasoning. In the first, children’s explanations are psychological, finalistic, and magical. During the second period, explanations are
artificialist and animistic, lack understanding of naturally occurring events, and focus on observable properties rather than the invisible components inherent in mechanisms, unlike adults. Therefore, he defines these two stages as pre-causal, and proposes that genuine causal understanding does not appear until about age of 7/8. In his proposal, Piaget mostly follows the Humean account and argues that while regularities are observable, causality is always unobservable and must be inferred using some set of theoretical presuppositions that are read into the observable features of events.

Several researchers replicated Piaget’s stage approach by also following the Humean account of causation, with tasks investigating children’s sensitivity to contiguity and regularity in their causal inference (see Siegler, 1976; Siegler & Liebert, 1974). Other studies employed tasks that included e.g. small toy cars designed to reveal children’s use of covariation in their causal analysis (Shultz & Mendelson, 1975). Later, Piaget’s views were challenged by contemporary research. In the non-Humean context, Bullock and Gelman (1979) explored preschoolers’ judgments about the action at a distance where connections between cause and effect were not necessary. Originally, this idea was proposed by Piaget (1930), as he stated that children could think when one pushes the pedals the wheels would turn. In another experimental design Bullock, Gelman, Baillargeon (1982) devised a mechanistic toy that a marble ball and a rolling light could trigger the Jack to pop up with a preferred time delay. The authors reported that preschoolers could judge the rolling light as a cause of Jack popping, with the condition that if they think that the marble is not the cause. When the temporal gap between marble rolling and Jack popping extended, pre-schoolers considered other possibilities, such as rolling light, which suggested that they have required competences for causal thinking with various types of events (see also Gelman & Spelke, 1981).

As discussed under the scientific thinking literature, the main contribution of these studies in the literature is that they pioneered the discussion on how mechanism knowledge influences causal reasoning, in particular when causal representations are not embodied covariations. Michotte’s (1946/1963) pioneering work on perceptual causality also shifted out from Humean view and emphasized the role of perceptual spatial-temporal properties in causation as discussed broadly in chapter three. Leslie and Keeble
(1987) also demonstrated that perceptual spatial-temporal elements are detectible from infancy. Various studies provided supporting evidence on these (Cohen & Oakes, 1993; Gelman, Bullock & Meck, 1980; Schlottmann, 1999; Schlottman et al., 2002; Scholl & Nakayama, 2002; and see also studies by Spelke, Phillips, Woodward, 1995; Gelman & Kremer, 1991; Woodward, 1998, 2009; White, 1995, 2014; Saxe, Tzelnic, Carey, 2007; Muentener & Bonawitz, 2017, on the importance of recognition of agents and their actions).

Two broad strands of work have been feeding into the causal cognition literature: causal learning and causal reasoning (Danks, 2014). Causal learning roughly corresponds to the acquisition of new causal beliefs, while causal reasoning puts more emphasis on procedural knowledge, and the ways existing representations change. Causal learning research concerns with the ways causal relations are induced from data, with limited space for other types of reasoning, such as spatial or spatial-temporal analysis. In this view, learning occurs from the patterns of repetitions or covariation, as a necessary condition to assess the strength of the causal relations. On the other hand, one of the prominent accounts of reasoning suggests that causal reasoning involves the construction of possibilities and the search for counterexamples, where reasoning is the source of knowledge about interactions among events and processes (Goldvarg & Johnson-Laird, 2001; however, see Lagnado, 2009, 2011, for a review on causal learning and reasoning being subject to similar constrains due to use of common cognitive resources, such as attentional resources or working memory).

Various modes of human inference have been studied within these two strands. For instance, Cheng and Holyoak’s (1985) study discusses a variety of knowledge structures such as syntactic versus pragmatic, which are obtained inductively from life experiences. Rescorla and Wagner’s (1972); Shanks and Dickinson’s (1987) works focus on associative accounts of learning (see also Sloman, 1996 for a review). Studies also show that humans can conceive of cause-effect relations from a variety of cues, such as statistical relations, temporal order, intervention, or prior knowledge (Lagnado et al. 2007). Research distinguishes in particular between two modes of access to causal knowledge: through observation of regularities, and experimentation/intervention (see
Gopnik et al., 2004; Danks, 2005; Lagnado & Sloman, 2004; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003; Waldmann & Hagmayer, 2005). Some minimal prior knowledge about the structure of the causal events is also required (Waldmann, 1996; Lagnado, 2011), but this knowledge grows through either observation or intervention, and is experientially driven.

The distinction between observation and intervention is consistent throughout causal learning studies. In particular, it is a focus of work on causal inference based on Bayes networks (Gopnik et al., 2004), employing top-down (Waldmann, 1996; Lagnado & Sloman, 2004) and bottom-up Bayesian approaches (Heckerman, Meek, & Cooper, 1999). Comparing the two, evidence suggests that learning through intervention is seen to be more efficient than observing, for the reason that (i) interveners can select and access to a variety of data, allowing them to choose the type of information they want to focus on; (ii) unlike observation, people need to make a decision on which intervention to apply; and (iii) in an intervention, causes precede their effects and a temporal cue is implicitly available to the reasoners (Lagnado & Sloman, 2004).

This thesis focuses on causal reasoning, rather than learning, and aims to provide further evidence for whether observation can provide advantage when the focus is not learning. There is a study indicates that when people make inferences about a known causal structure, they represent interventions distinctly from observations (Lagnado & Sloman, 2002), whereas when they arrive at causal structure inductively, they do not separate observation and intervention explicitly, but instead focus on temporal cues (see Lagnado & Sloman, 2004; Lagnado et al., 2007). We can of course talk about different forms of observation, such as data driven or online. The point here is that while temporal information is a stable cue in intervention, it may not be a prominent factor for data driven observation (Lagnado & Sloman, 2004). It is unknown whether spatial and temporal information is a stable cue when individuals directly witness continuous causal processes in an online fashion. Two reasons constitute our lack of knowledge. First, past research has been dominated by the studies examining how people induce causal relations from patterns of covariation (Lagnado, 2011), such as inferring causal structure from statistical information or single causal events (Cheng 1997; Griffiths & Tenenbaum,
2003; Hagmayer & Waldmann, 2000; Jenkings & Ward, 1965; Shanks, 1995), or inference multiple cause-effect relationships from causal networks (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). Second, the role of spatial-temporal analysis in causal cognition is highly neglected. In that context, no study has yet investigated the kinds to access to causal knowledge of natural phenomena through observation, how this develops, or in particular, whether observation of natural phenomena mediates access to causal mechanism effectively.

Cheng (1997) highlighted that causal reasoning literature has been dominated by two approaches. The first is covariation, which can be traced back to Hume, and is also the source of the contingency model, touching the basis of the rate of the strength/probability of the causal relations (see also Jenkings & Ward, 1965). The insufficiency of the covariation approach encouraged researchers to ask how to move from covariation to causation. The alternative view, the power approach, therefore traces its roots back to I. Kant. This view relied on the proposal that people have a priori knowledge about events and their causes, but alterations take place due to conformity with the connection of cause and effect. Note that the studies by Ahn, Kalish, Medin, Gelman (1995); Bullock, Gelman, Baillargeon (1982); and Shultz (1982) can be seen as Kantian, since they showed that people infer a cause of an effect when they perceive or know the specific generative source, or the causal mechanism, linking the candidate cause to its effect.

Consequently, various specific theories nourished causal reasoning literature, and those theories can be classified with regards to the type of causal relations they represent. According to Woodward (2016) theories of causation can be classified under at least three categories: regularity, counterfactual, and process theories. The guiding idea of the first, the regularity theories, is that causal claims impose a regularity linking a cause to its effect in a deterministic way. The second idea is that causal claims are closely linked to counterfactual relations, which can be understood as the account of possible worlds (if A were the case, B would be the case) as a function of our intuitive judgments about causal dependence. He adds that the sciences are full of counterfactual claims. The third class, causal process theories are considered together as a ‘difference-making’ account of causation, with the idea that a cause makes a difference to its effect. As outlined briefly
in the introduction, the process view of causation has largely been neglected in psychology literature.

Unlike Woodward’s classification, Waldman (2017) defines three categories for causal theories: the dependency, the process, and the disposition frameworks. The dependency framework involves most of the above-mentioned psychological theories (e.g. from associative, causal learning, to Bayesian inference). The core idea of the dependency framework is that the variable C is a cause of its effect E if E depends upon C. Most theories within the dependency framework concern with the modelling of causal relations and how people infer about causes and effects as discussed before. The disposition framework concerns with the interrelations between the objects involved in causal interactions. For instance, in the analysis of the aspirin and removal of a headache, this approach proposes that the dependency relations are secondary, but the interplay between the aspirin and human body should be analyzed, as the disposition to the influence could be circumstances specific. Within the process theories, event representations are abstracted over the causal processes. This thesis relies upon the process view of causation, and value causal reasoning within the physical domain, as discussed further.

1.4.1 Contributions to the causal reasoning literature

This thesis contributes to the causal reasoning literature in at least two novel ways. Firstly, various researchers sought to characterize the ways by which people establish causal models or representations from covariation information. In this context, the role of spatial-temporal cognition in causal reasoning has been highly neglected. Although the term spatiotemporal or spatial-temporal has been used since Hume, surprisingly the underlying cognitive processes of this specific cognitive domain have never been studied. Instead, the theoretical approach has surprisingly been received more interests from computer scientists (see an edited book by Stock, 1997). This is highly crucial to understanding what makes an entity, event, and a process perceptual, rather than inferential; how perceptual primitives influence the grasp of causal processes. This thesis postulates that perceptual primitives involve (i) qualitative, such as colour, texture, smell, topological structures; (ii) quantitative, such as length, width, depth in the Euclidian fashion; and (iii) temporal properties, such as speed, succession, directionality, which are
embedded in three/four dimensional experienced space-time. For instance, imagine that one is observing an apple rolling/falling, in which the temporal onsets and offsets are within the available environment. While motionless, the qualitative and quantitative features of the apple can be afforded. Once, it starts rolling/falling, various features move together, requiring perceptual primitives to be bound across space and time, and enabling online registrations. Employing perceptual, analytical, and imagery-based kinds of spatial-temporal tasks, this thesis aims to provide a behavioural analysis of how perceptual primitives can be mentally abstracted; integrated by higher level intellectual faculties, such as conceptual reasoning, or imagination; and how this abstraction/integration relates to reasoning about causal processes.

Perceptual primitives are also fundamental to natural phenomena, such as when an object sinks in water, moves on the floor, or flies in the air. However, the capacity to utilise the primitives and the capacity to reason about causal processes might be independent analyses, as hypothesized in the third chapter. For instance, the ability to observe a ball travelling in the air; guessing its landing position; and reasoning about the cause-effect relations (ball throwing-falling) seem to require various competences. Children may be sensitive to the length of the demonstrations, or as McCloskey studied, this may rely primarily on prior knowledge, or alternatively prior knowledge and the primitives should be combined for reasoning. To answer these questions, the methodology involved single demonstrations of natural phenomena. Due to the use of natural phenomena in causal reasoning literature being very rare, in particular with regard to sinking, dissolving, solution, those three phenomena were materialized with substances prevalent in daily life. Baring in mind previous research, informing us that extracting new information is possible through a single learning episode, and even young children can change their representations or mental models about phenomena through single demonstrations (see Shank, 2004; White, 2014, for a discussion). This is an alternative view to the regularity approach and highly cited covariation literature, though the link between statistical thinking and spatial-temporal analysis is also investigated in the third and fourth chapters.

The second novel contribution of this thesis to the literature is that causal reasoning tasks in both developmental and adult literature typically involve simple machines, such as
physical or virtual apparatus with distinct components yielding a well-defined, segmented sequence of events. Such structure helps operationalize what may be cause, mediating mechanism and effect, allowing experimental manipulation of what can be observed and has to be inferred. Some forms utilize contingencies or statistical information to allow reasoners to infer underlying causal structure/strength that generates patterns. Other adult studies involve descriptions of causal sequences (e.g. Buchsbaum et al., 2015; Talmy, 1988; White, 2014), which help structure events that in real life may be more continuous and less easy to segment. Some paradigms in the tradition of Michotte (1946/1963; Schlottmann et al., 2006) involve reasoning about brief events, mostly collisions, which also provide natural segmentation. This schema of causality between distinct, clearly segmented events is common to the vast majority of psychological studies.

This can also explain the gap why process view of causation is largely neglected in causal reasoning literature: we did not have an experimental methodology to explore the cognition of causal processes, in particular, when natural systems typically involve temporally more extended continuous processes, as when an object sinks, dissolves or absorbs water. Here there are no distinct cause and effect events; a standard causality-as-events approach would imply that the hand letting go of the stone is the event that causes its sinking, but this is intuitively not the right cause and would be more appropriately regarded as an enabling condition. Salmon (1984) distinguished causality in such continuous processes from causal interactions (see also Kitcher, 1989). Considering this distinction, sinking involves an interaction of a stone subject to gravity with water pushing upwards, where density is a hidden feature of the object that needs to be observed from the interaction with water. Note that the process theories such as Salmon (1984) and Dowe (2000) distinguish between causal processes and causal interactions. In particular, in Salmon’s version a causal process refers to a physical process, with difference making treatments of causation. For instance, any movement through space transmits energy and momentum or some other conserved quantity in a spatiotemporally continuous way. Therefore, while processes can – and perhaps need to – be considered as chains of interactions, no interaction is actually perceived – one sees the stone sink, but must think about the role of the water. Observers therefore do not see the sinking as either caused by a distinct event via a mediating mechanism, or as an interaction between stone and water; the phenomenal experience presented by the causal sequence is of a
uniform, unsegmented process caused by a continuously operating underlying mechanism.

Reasoning about causality-as-events, interactions and causality-as-processes all aim to distinguish true from spurious or pseudo-causality by appeal to an underlying mechanism. In each case, people must reason beyond what may be observed at the perceptual surface. However, the distinct components of machines typically reveal the operation of the underlying mechanism more clearly, while it may require more effort to go beyond the perceptual surface of continuous processes. The argument therefore is that reasoning about causal mechanisms underlying the simple physical processes encountered in elementary science is a step up from causal reasoning about the toy-machine systems previously studied, and investigate here children’s ability to extend their causal thinking to these.

The notion of *mechanism* has a unique importance across the thesis. It needs to be elaborated, as there is no readily available or accepted conception about it. For instance, in philosophy, one of the views states that causation and mechanisms are potentially interdefined, stating that mechanisms are sequences of interconnected events (see Glennan, 2002, for a review). Another view analyses mechanisms as complex systems that involve parts and processes (e.g. Bechtel & Abrahamsen, 2005; Salmon, 1984; Machamer, Darden, Craven, 2000). In psychology, Piagetian framework follows a mechanistic view and refers to mechanical systems. In this tradition, the mechanism makes a bicycle to work. Johnson and Ahn (2017) define mechanisms as systems of visible and invisible characteristics interacting systematically, where the same effects are produced by the same causes. Park and Sloman (2013) refer to the set of causes, enablers, disablers, and preventers that are involved in producing an effect, including how they unfold over time. Lombrozo (2006; see also Lombrozo & Carey, 2006) states that mechanisms have a privileged relationship to explanations; they do not simply identify causes, but illustrate how the cause brought on the effect in a mechanism.

In this thesis the term mechanism refers to the notion of explanation seeking for *how a phenomenon can be understood* or *how a process works*, as commonly used in scientific
fields such as biology, neurology, chemistry, or physics. This approach follows Salmon’s and Machamer et al.’s definitions explained below. From a dualistic perspective, mechanisms are composed of both entities (properties) and activities (change producers) that enable a phenomenon to happen. Entities are regularly located, structured, and oriented, and the activities engage in a temporal order, rate, and duration (see Machamer, Darden, Craven, 2000, for the dualistic definition of mechanisms). In Salmon’s (1984, 1998) view of causation, mechanisms are dynamic, composed of processes, which refers to a phenomenon that exhibits consistency over time; and interactions, which refers to spatial and temporal intersections that change processes persistently. The regularities therefore are not accidental. This view corresponds to the causal realist theory, which sees causal mechanisms dissimilar to the Humean account. As discussed earlier, the Humean account derives causation from empirical regularities. Works inspired by the Kantian account derive causation from a priori concepts or categorizations of understanding. For instance, the causal power implies the pathway that link A to B causally (Salmon, 1984; Miller, 1987, for a review). According to this account, causal explanations are derived from law-like structures and their theoretical principles are universally reliable.

For the first time, in applying this paradigm to cognition of physical processes, this thesis aims to explore the developmental courses, particularly how children’s reasoning about how a process works and how this reasoning develops across primary years. In developmental psychology, Bullock et al.’s (1982), and Schultz’s (1982, 1986) studies can also be classified in this category, as these studies focused on the causal judgments rather than looking at associations or regularities. I discuss these studies in more detail in the chapters, but refer them here to highlight the difference between those studies and the focus of this thesis: the previous studies put more emphasis on children’s inferences of causal mechanism. Although Piaget’s (1930, 1970/1972, 1969/2006, 1971/1974), and Bullock et al.’s (1982) studies used the term spatial-temporal thinking/reasoning, the cognitive processes involved in this kind of analysis were highly neglected. Moreover, this strand of work can be classified within the dependency framework classified by Waldman (2017), where causal processes were out of focus. More specifically, in Bullock et al.’s studies causal relations were presented with machines as part of the event
demonstrations. Children needed to think about the mechanisms in Piagetian contexts; thinking about processes was not the concern.

There is also other work showing that pre-schoolers engage in inferences about mechanism in clearly structured causal systems, typically simple machines, provided they have prior knowledge or experience of these (Buchanan & Sobel, 2011; Bullock, Gelman & Baillargeon, 1982; Schlottmann, 1999; Shultz, 1982). Buchanan and Sobel (2011), for instance, showed children that pressing one of two buttons made a light go on; the causal button had a sticker and was connected to the light by a wire or had a battery inside. Four-year-olds predicted that if the wire was switched, the other button would make the light come on; even 3-year-olds predicted this if the battery was switched. Neither age thought switching stickers would affect the outcome, indicating prior knowledge of what was relevant to the mechanism, and concern with how this led from cause to effect beyond what was observed.

Although such demonstrations show pre-school children are able to engage in forms of mechanism-based reasoning, similarly, the tasks typically consider the effects of mechanism-related variables on outcome. Children were not asked about, nor understood, presumably, the processes by which these variables exerted their effects; for instance, the operation of electrical circuits or chemical reactions inside the battery. Little is known about how this level of understanding develops, or how children go on to reason causally about less clearly structured or never fully observed processes in the natural world, although they encounter these in much of elementary school science. There are related literatures, such as work on children’s intuitions regarding physical phenomena (see Wilkening & Cacchione, 2011), but the tasks used in this are designed to elicit pre-existing and implicit knowledge of causal connections, not explicit reasoning about processes to further such knowledge.

Overall, mechanisms specify causal information through generating processes. The regularity approach instead emphasizes a general process involving the analysis of covariation between factors and effects, not the ‘how’s (Ahn et al., 1995). From this distinction, in this thesis the notion of process and mechanism are seen as accompanying
concepts, rather than fundamentally distinct. In this paradigm, spatial and temporal properties represent accessible and observable qualities of causal processes. However, they can still highlight non-obvious and abstract factors that intervene in such processes. For instance, even in the most abstract forms, such as observation of interactions, transformations, or growth, spatial-temporal properties can cue causal relations, in particular in single cases. Although scientific minds seek for further evidence or regularities to infer mechanisms (how a process works), spatial and temporal properties still need to be considered for the first and following observations for each case.

Consequently, studies on causal mechanism typically ask how mechanisms enable causes to produce their effect(s). This brings into mind a system of abstract variables, consistently tying a cause with its effect(s), but each variable has its visible properties spanning over space and time. Although causal mechanisms determine law-like aspects of causal processes in a more deterministic way (gravity causes falling objects to accelerate); neither processes, nor mechanisms are easily observable, but once mechanism is known (gravity causes acceleration), processes can be understood/manipulated with the aid of spatial and temporal properties (falling objects).

1.5 Explanation and causal reasoning with words

It is the interplay between experiences, representations, and reasoning that drives humans to create various intelligent ways to make their knowledge and representations available to others. That could be a formula, body movement, a diagram, or a sentence. Among different theoretical approaches to ‘what an explanation is’ (see e.g. Pitt, 1988; Hempel, 1965; Hempel & Oppenheim, 1948), knowledge can be in explicit or implicit forms. And in psychology, various explicit forms are composed under the umbrella term of explanation to refer to the forms that permit the deliberate use of reasoning skills and the production of new representations. For instance, there are various categorizations of explanations, from everyday to nomological (law-like), or causal. Last two kinds depart radically from everyday explanations, as they suggest mechanisms whereby people can recruit them on the basis of one or multiple trials (Keil, 2006).
Salmon (1984) distinguishes between the two kinds of explicit knowledge: description (knowledge of what) and explanation (knowledge of why). While descriptive knowledge is a portrayal of observed events, it is explanation that grows out of causal connections, fertilize the understanding, comprehension and enlightenment, which also enables us to anticipate future outcomes, predict the consequences of events and test our understanding of the regularities involved. Lombrozo and Vasilyeva (2017) also highlight that not all explanations are causal, neither all causes are explanatory. For instance, mathematics typically involves various non-causal structures. People can engage in explanations to justify an action, or in the service of aesthetic pleasure, or predict the outcomes of future events (Keil, 2006). However, in science, explanations typically refer to causal relations, or causal factors that permeate in particular circumstances. Considering an explanation for climate change, causation becomes intimately related to explanation, as it often reveals the causes and effects or e.g. how climate change occurs over a certain period of time.

Developmental research of causal explanation was initiated by the work of Piaget (1959, 1930/1960), who proposed that children’s failure in explaining scientific phenomena was not due to a lack of observation or ignorance of a law of nature, but lack of operational thinking and the ability to master the use of various transitive verbs, which both improved with age over formal operational stage. Piaget’s proposal was shaped by his methodology, which mostly contained open-ended questions about natural phenomena like ‘where the sun came from’. The Piagetian approach put concrete barriers between the stages, and his tasks arguably involved complex phenomena. Therefore, he proposed a delayed understanding, where for instance young children’s explanations were defined as pre-causal due to their inability to distinguish between psychological activity and physical mechanism. Since then the theme of developmental change in children’s explanation has received interests, but current studies employed more child friendly tasks.

From a non-Piagetian perspective, Bullock et al.’s (1982) above mentioned studies revealed that children’s explanations do not reflect the level of sophistication of their reasoning about causal mechanisms. This becomes more salient when the tasks contain
nonverbal protocols that reduce the influence of linguistic competences in task performance. Gelman & Kremer’s (1991) tasks combined what and why questions and required children to distinguish natural phenomena from human-made artefacts. This work documents that children can give reasonable explanations of simple events from the preschool age (e.g. why leaves turn brown), but the sophistication and accuracy in their explanations increase with age. Wellman and Gelman’s (1998) study supported this view and proposed that children’s content knowledge of physical, biological, and psychological phenomena develop very early on. With the development of language, children start showing causal-explanatory abilities in greater or lesser depth depending on the domain and the developmental changes.

Legare, Gelman, and Wellman (2010) investigated whether preschoolers make use of causal knowledge in their explanations. Children were shown two equivalent events, one in accord with prior knowledge and the other not. For the former, children were shown a set of novel light box devices. The experimenter activated the light by placing an object on top. Children were briefly taught about categories of objects that turn the box on and off (stoppers, starters, do-nothings). After training, children were introduced scenarios requiring them to explain whether the new object looked like a starter, stopper, or a do-nothing. The authors asked children why did that happen, and looked at children’s explanations –what they explained and how they did so- (e.g. whether they considered category membership). The authors found that novel event demonstrations, where children had no prior knowledge, powerfully triggered children’s explanations. Moreover, the authors focused more on explaining events, which were inconsistent with their prior knowledge, to clarify whether they had readily available language for an unknown or an unexpected outcome. The message was that by the age of three, children could use causal knowledge to make predictions and construct mental models to anticipate regularities. The result was in consistent with Shultz’s (1982) work.

The non-Piagetian routes underlined that the apparent quality of explanation might reflect limited language not limited reasoning: although children’s understanding of natural causal events is often sophisticated, they may perform poorly compared to adults at expressing and assessing their own causal knowledge. For this reason, developmental
psychologists in recent years have attempted to capture the dissociation between verbal and nonverbal reasoning, each requiring different domains of knowledge (see e.g. Legare & Clegg, 2014; Legare & Lombozo, 2014; Mandler, 2004), such as expressive vocabulary, visual, spatial-temporal, or symbolic knowledge.

Other studies show that children are active causal explanation-seekers, but they continuously change conceptual structures about their naïve theories and renew the types of processing over the development. This framework theorizes that domain general executive function abilities are involved in conceptual change processes as because conceptual change requires the inhibition of previous misconceptions (e.g. Carey, 1985; Zaitchik, Solomon, Tardiff & Bascandziev, 2016; Vosniadou, 2014; Gopnik & Meltzoff, 1997; Legare, 2014). The importance of this strand of work is very little for the scope of this thesis, as because this theory fails to explain how actually children reason about novel causal relations when prior belief/knowledge is not primary. The second point is that this thesis does not concern with conceptual change or children’s prior misconceptions.

Recent studies in psychology address the role of explanation in cognition in different ways. For instance, the types of explanations are analyzed as being intentional, causal, and functional as discussed below (see e.g. Brewer, Chinn, Samarapungavan, 1998; Lombrozo, 2007, 2016). The notion of intentional explanation refers to beliefs and desires (Bloom & German, 2000). Causal explanation is seen as an account that provides consistent conceptual narration about the phenomena. Regarding this narration, recent psychological approaches compare children’s explanations with scientists’ and explore the criteria that are used. Samarapungavan (1992) shows that children can use most meta-conceptual criteria that also scientists use. For instance, when it comes to evaluating theoretical alternatives, children consider empirical and logical consistency supporting their prior beliefs. Brewer et al. (1998) highlights the large overlaps between everyday and scientific explanations, and also proposes that children use the same evaluation criteria in their explanations as adults and scientists. Some of the criteria are empirical accuracy, scope, consistency, simplicity, and plausibility. According to Brewer at al., children can utilize most of these common forms of explanatory frameworks also used by
scientists, except that they differ in the use of formal and mathematical accounts that certify precision. This strand of work seems to contradict with most studies discussed in the scientific reasoning literature, in particular the ones that suggest late development in metacognitive abilities (i.e. Dean & Kuhn, 2007; Kuhn, 2007; Kuhn et al., 2008).

Studies on the function of children’s explanations show that not only a single explanation is evaluated in isolation, but also various competing explanations are embedded in a network together with other related information. This implies that a selection process is involved in explanatory competences. Lombozo (2007, 2016) argues that simplicity is one of the most crucial factors for evaluating and selecting among other explanations. Even 6-year-olds prefer simpler explanations, as simpler ones invoke fever causes (Bonawitz & Lombozo, 2012). Others argue that when the function is considered, children are teleological. For instance, from a teleological perspective a child can propose that rain exist for plants to grow (see e.g. Kelemen, 1999). Lombozo and Carey (2006) suggest that selective application of teleological explanation has a deep commitment to the causal structure of the world: teleological explanations are interpreted causally in the condition that when the function invoked, the explanation consistently plays a causal role. However, in the context of scientific thinking Keil (2006) distinguishes between the teleological and causal explanations: the latter refers to relations in at least four distinct ways, such as common cause, common effect, causal chains, and causal homeostasis (why sets of things endure the stability of the mechanism). On the other hand, the teleological explanations slightly concern with intentional desires.

The complex nature of causal explanations requires also counterfactuals to be explored (if A had not occurred, then B would/not have occurred). Although research in this strand is limited, developmental studies point that even young children use counterfactual thinking in causal reasoning (German & Nichols, 2003; Haris, German, Mills, 1996). However, this strand of work is not closely relevant to the scope of this thesis, as children’s capacity for such counterfactual thinking is not the focus here. Plus, counterfactuals rely primarily on AB type distinct event approach, in which participants need to think through a counterfactual sequence where two or more antecedents would
or not play role in differentiated effects. Moreover, with its highly hypothetical nature, counterfactual thinking largely differs from the ability to analyze spatial-temporal information from experienced space, as the latter relies primarily on observation and utilizes of spatial-temporal representations in causal analysis, while the former requires one to imagine worlds that do not exist.

A more convenient explanatory framework for the scope of this thesis underlines the two means by which people can identify causal relations: induction (inference under uncertainty) and abduction (inference to the best explanation) (Ahn & Kalish, 2000). An inductive inference takes into account the characteristics of observed samples to conclude the distribution of novel causal relations in larger population. An abductive inference takes into account a specific case and seeks for the best explanation by rules or observations (why X, but not Z is the best explanation of Y). The assumption of abductive reasoning is that our knowledge about causal mechanism is used to determine the best explanation for a phenomenon, like a theory explains evidence. Therefore, it is compatible with mechanism approach (Josephson & Josephson, 1994). Within this view the explanation process is conceptualized through three stages: generation, criticism, and acceptance of explanatory hypotheses. The explanation process aims to explain why a theory explains evidence better than the alternatives. In Peirce’s (1955) words, “Given an observation d and the knowledge that h causes d, it is an abduction to conclude that h occurred” (p.14). According to this view, even simple explanations are typically composite; all develop through the same process, with the three parts (generation, criticism, acceptance) holding together in various ways.

The fact that literature on language development and its determinants is well established (see e.g. Bishop et al., 2016; Carey, 2009; Karmiloff-Smith, 1992, for further reading), prompting us to consider that drawing an appropriate causal explanation must be a complex and multifaceted process, requiring prior knowledge to be synthesized with the concepts and their relations. A rigid counter argument, taking into account the role of language, may suggest that children’s explanations are the product of their language alone. Thus, their explanations package up all aspects of their understanding and communication. There are some studies providing contradictory evidence against this
rigid argument. For instance, Diesel’s (2004) study showed that even toddlers, around age of three, could connect distinct events and explain the causal relations. Similarly, research looking at young children’s conversations with their caregivers highlights that children start providing causal explanations around age of 2-3, and this increases in frequency across development (Callanan & Oakes, 1992; Wellman, Hickling, Schult, 1997). By the age of 5 children seem to be able to answer ‘why’ questions and complete ‘because’ sentences from event demonstrations (Donaldson, 1986). In this thesis children’s verbal reports are analysed to elaborate how they reason about observed causal processes. Examining children’s explanations is quite common (e.g. Piaget, 1929; Wellman, Hickling, and Schult, 1997), but the role of nonverbal dimension in reasoning about causal processes was also addressed to provide insight into how this kind of thinking develops towards more demanding forms. For that reason, both verbal and nonverbal dimensions were included in causal task protocols, and also standardized verbal and nonverbal measures were included in the battery as direct tests of how far children’s explanations of causal processes are specifically a function of verbal or nonverbal ability across development.

However, the counter view still warns us to be cautious and not to take into account various forms of language production affecting children’s competences. In order to avoid possible confounding effects, various forms of language ability were out of focus in the scoring systems used here, such as individual differences in word learning, the development of language or vocabulary comprehension, the development of grammar knowledge, language acquisition, or the complexity of the sentence structures. Instead, the focus was on children’s ability to predict, describe, and infer single causal processes, connect observables and variants with an awareness of unseen mechanism. The scoring system therefore considered what already children intuitively knew about the three phenomena (prior knowledge from prediction), what happened when they were asked to observe and evaluate the status/amount of the information drawn from the observations of causal processes. In this design, children’s explanations did not rely solely on prior knowledge. Verbal reports were analysed in a structured fashion, separating explanation - as a measure of ability to reason beyond what was observed- from prediction based on prior knowledge and description of current observation. The measure of explanation was designed to distinguish between children who could only identify simple factors at work
in the observed phenomena (i.e. weight), those who recognised that variation in these factors is associated with differences in speed of process (i.e. who treat them as variables), and those who were able to move beyond variables per se, to think about the underlying mechanisms connecting variables to outcomes. The measures of prior knowledge and description made it possible to assess how far level of explanation was a function of existing knowledge versus current observation.

Children’s explanations about three phenomena were investigated via two main questions ‘why X occurs in that way’ and ‘do you think there may be another reason for that’. In this design, participants are potentially expected to take a diagnostic approach to come up with the best explanation for the causal outcome they witnessed (similar to the abductive reasoning approach explained above). Taking diagnostic fashion as abduction, the classes of variants needed to be considered, just like how a doctor diagnoses a stomach infection: children needed to organize the alternative thoughts, then describe and evaluate the observables (evidence) against their thoughts/hypothesis, and come up with their best explanation. The power of this approach relies on its strength in controlling for the explanation process. For instance, in the case of an inadequate answer, one can elaborate on whether hypotheses are incorrectly judged, or observables are inadequately utilized, or diagnostic conclusion is not sufficient (see Ahn & Kalish, 2000; Josephson & Josephson, 1994, for a review). Children’s initial answers were also followed up to minimize verbal demands and response ambiguities by posing further structured questions and by supporting their expression of knowledge through careful probing.

The criteria for judging a good explanation were not the amount of the sentences or their theoretical background. Instead, the methodology is guided by the appropriateness of the explanation in such a way that, for instance, if a 5-year-old says ‘the stone has more stuff in it and it is compact, but the grape is not. That is why it sank faster’, this is count as corresponding to the mechanism level explanation without use of the word density. Understanding and coordination of observables with the variants, rather than word knowledge is of interest. Of course, one can argue that children’s language production to be noticeably different with age, as a consequence of the improvement in their word and domain knowledge. However, the relaxed scoring system here aims to not to disqualify
young children, but instead to take into account their reasoning and use of variables to think about mechanisms. The difference between previously employed methods and the method used in this thesis therefore is that all studies here employed the real phenomena; children were demonstrated the real objects with direct comparisons, and good amount of care to observation of the causal processes was taken.
Chapter 2

The way of information presentation may affect children’s ability to predict, describe and explain continuous causal processes

2.1 Introduction

Does children’s ability to infer cause-effect relationships depend on how representations are acquired and organised? The chapter aims to answer this question. It is unknown whether children can infer causal relations, particularly causal mechanisms involved, from direct observation of natural causal processes; and whether their capacity to do this is affected by how information is presented. Research highlights that the temporal dynamics of information presentation matter for event representation and category learning: changing the way in which information is presented and observed changes what is learned (Carvalho & Goldstone, 2015a; Schyns, Goldstone & Thibaut, 1998) due to effects of (i) familiarity (repeated experience), (ii) the type of relation between objects (degree of similarity) (see e.g. Mandler, 1986), and (iii) the sequence in which they are encountered (combined versus distributed). However, the role of temporal dynamics of information presentation in causal inference is unknown. By taking into account age related changes this chapter aims to test whether children’s causal inference is affected by information presentation, as hypothesized.

In total 156 children (5-to-11-year-olds) saw two contrasting instances of three natural phenomena (sinking, absorption and solution) under one of three conditions: paired-contrasting, sequential, and jumbled. For each condition, children’s responses were scored based on their prediction from prior knowledge, description, and explanation of the mechanisms involved. Results showed that (1) performance on all measures improved systematically with age; (2) there were no differences between conditions in overall performance, but inferences following the first instance in the sequential and jumbled conditions were significantly poorer than final inferences: availability of the contrasting example played role on this; (3) descriptions were related to inferences: accurate
observation was also crucial to understanding of causal processes; (4) prior knowledge brought into the tasks had at most a modest influence on performance; (5) inference of mechanism was comparatively rare in all conditions, and lagged behind the ability to predict and describe observations; (6) general nonverbal ability had greater impact on performance in the paired-contrasting condition, especially on mechanism inference, whilst verbal ability was dominant in the two other conditions.

The findings highlighted that children could infer causal relations directly from observation of natural phenomena, but their capacity to infer mechanisms remained immature, although highly dependent on information presentation. Spacing between observations appears to permit representations to be encoded verbally, but simultaneous presentation may promote perceptual awareness of the spatial-temporal characteristics of causal processes, providing a better stepping stone to mechanism inference. This chapter will discuss these broadly by taking into account two strands of research: the role of familiarity, and information presentation literatures.

2.2 The role of familiarity in accessing causal knowledge

The most commonly proposed reason why we engage in causal reasoning is that it enables us to predict similar events in the future (Keil, 2006). Even young children engage in inferring causal relations (Gopnik & Glymour, 2004). Evidently, it leverages access to much deeper comprehension, a crucial property of scientific thinking. The question is that whether children’s improvement in causal reasoning are attributable to the previous experience with phenomena, past or current familiarity? Early studies answered this question in a more conservative way. For instance, Berzonsky (1971) used three measures to investigate the effect of familiarity on children’s causal explanations: an interview questionnaire dealing with familiar, remote, and malfunction of objects (e.g. what makes a clock tick? What makes the wind blow? What makes tires go flat?); another two measures involved real demonstrations requiring children to predict and explain what had happened and why (demonstrating two objects and asking which one would raise the water level and why). He found that children’s experience and familiarity with the objects and various phenomena determined the type of causal explanations given.
The question ‘how people come to know the causes of events’ has been the subject of various contemporary researches as discussed in the first chapter. Similar to the scientific reasoning literature, the discussions on causal inference divided into two main components. One was concerned with how an acquired causal relation is first induced, and another considered how prior (domain-specific) knowledge influences causal judgments. Various experiments justified the notion that the causal induction component operates separately from prior knowledge (see Cheng, 1997; Cheng & Novick, 1990, for a review). The basis of causal induction, presenting formal analyses of how an agent learns about causal relationships, was studied by several models to give a better account of human judgments [e.g. see R-W model (Rescorla & Wagner, 1972), power PC theory (Cheng, 1997; Cheng & Novick, 1990), Bayesian inference models (Griffiths & Tenenbaum, 2005, 2009), causal models (Gopnik et al., 2004), probability based theories (Sloman & Lagnado, 2005) experimentation/intervention (Danks, 2005; Lagnado & Sloman, 2004), associative account for causal learning (Shanks & Dickinson, 1988), different forms of inference such as hypothesizing (Waldman; 1996), causal heuristics (e.g. Kahneman & Tversky, 1982) and counterfactual thinking (e.g. German, 1999)].

The role of prior knowledge became more prominent when causal structure and strength was distinguished -the former asks whether causal relations exists, whilst the latter focuses on how strong that relation might be- (Griffiths & Tenenbaum, 2005; Schulz, Kushnir, & Gopnik, 2007). For instance, to distinguish between causal structure and strength Gopnik and Sobel (2000) designed three experiments to explore young children’s understanding of causal powers. The authors set up a novel causal power of objects (whether or not the real objects – e.g. blickets, tibs – light up a particular machine –blicket detector–) and examined whether children use the novel information about the causal power of the objects to categorize them. In the control condition, the authors also explored whether children performed similar categorizations when there were associations (not causal) between the objects and the machine. Across the experiments children as young as three years old were able to categorize and name the objects based on the causal powers. Moreover, without age effect children could use the names to predict causal powers even when they conflicted with perceptual properties of the objects (e.g. shape, color). This study suggested that children do not just infer correlational patterns from their past experiences, but that they can infer and explain novel causal
relations from events they had not previously witnessed. Further studies (e.g. Gopnik et al., 2004; Kushnir et al., 2010; Schulz et al., 2008) explored how children understood principles, such as covariation, to infer causal relations when they have a chance to intervene on the causal system, even inferred unobserved common cause if two variables were correlated.

Rotttman et al. (2013) state that most real-world causal phenomena unfold over time, and therefore observing temporal scenarios is crucial to capture causal relations changing from one state to the next. However, studies of associative learning theory point that people fail to identify causal relations when the temporal lag between cause and effect is longer than two seconds (see Shanks et al., 1989). Buehner and May’s (2010) study presented a computational approach and highlighted that when the reasoners have prior knowledge of delay (participants’ expectations) their causal judgments change. Temporal delay effect becomes a weak evidence for inference then. We need to bear in mind that although these studies largely employed perceptual spatial-temporal elements, as typically studied in the Michottean framework, the findings suggest that temporal delay may negatively effect the ability to identify causal relations during observations, but enhancing the predictability of the outcome can change this pattern. If one has a prior knowledge about a phenomenon, this builds some sort of causal expectation beyond temporal scenarios. Causal reasoning may necessitate or is affected by some level of prior knowledge then.

On the other hand, there is some evidence suggesting that familiarity is not a prerequisite for inferring causal structure. For instance, in Bauer and Mandler’s (1989) study testing immediate and delayed recall of simple event sequences, two-year-olds organized their representations around temporal order, but when novel sequences of events lacked causal relations, the organisation of new representations was shallow. In particular, children encoded the temporal order of familiar sequences (i.e. giving a teddy bear a bath) and novel-causal sequences (i.e. a mechanism on a board causing a frog to jump) more reliably than novel-arbitrary sequences (i.e. drawing a picture on a board). This suggests that causal relations in even novel information are encoded differently from very early
on. Supporting that, the temporal order of novel causal sequences was reliably encoded even when the connection between the elements was interrupted.

However, even if children are sensitive particularly to the temporal order of causal sequences from their first experience, their representations of these typically become more elaborated and better organised with repeated encounters, especially when their representations are temporally structured (see Fivus & Slackman, 1986). For instance, McCormack and Hoerl (1999) report that children remember sequences relative order of familiar events located in time. Considering the effects of talking about on-going and past events Fivush et al. (1997) show that children recall past experiences better, but this interacts with the role of language, which may help to organize the representations. Unlike Bauer and Mandler’s argument, these studies (see also Fivush et al., 1992) support the idea that both children and adults generally exhibit better memory for events that are repeated than for events experienced a single time.

In the present study, the concern is not event knowledge in itself, as derived from repeated experiences, which potentially helps children to predict outcomes from their previous representations, nor memory effects on learning, or enhancing children’s event recall. Any research design focused on learning ensures participants receive feedback on their conclusions/choices or else undergo a habituation/training schedule, but the tasks used in this study did not include any learning phase, feedback, or intervention. The aim instead is to explore how causal relations can be induced from on-time observations via temporally extended processes. In particular, this study aims to explore what helps children to go beyond given information for real world phenomena to make inferences about the nature of the causal connection – the underlying process or mechanism.

There are counterarguments in the literature, proposing that children’s learning or understanding improves with direct instruction or teacher-centered methods (Klahr and Nigam, 2004). In two meta-analyses using a sample of 164 studies, Alfieri et al. (2011) also found that unassisted discovery did not benefit learners, whereas feedback, examples, scaffolding, and explanations did. This study acknowledges these proposals, and agrees that children are unlikely to discover advanced scientific topics, such as
geometry, algebra, chemistry, programming etc. Therefore the argument on children’s need of instruction is plausible (see also Anderson, Corbett, Koedinger, & Pelletier, 1995; Klahr & Carver, 1988). However, even if instruction matter, models of conceptual growth still assume a corresponding need for active engagement and mental construction on the part of the learner and discovery methods, allowing children to construct their own ideas of scientific phenomena (Tolmie & Dündar-Coecke, in press). This makes it important to examine how inference occurs when children are not given target information, but they find their answers independently from a range of presented experiences in more fluid fashion. One can assume that this is a more advanced form of thinking, as causal processes necessarily involve invisible as well as visible factors, and humans seek for understanding when it appears, and also whether specific forms of organisation of observation of causal sequences facilitate it. The data here aim to provide evidence for this matter.

2.3 Effects of information presentation

It is unknown whether the demonstration of similar or different objects, or the temporal sequence in which they are presented changes causal inference. Although the focus is not learning, relevant findings were provided by studies of category learning. In this strand, research documents that comparisons benefit object recognition and discrimination. Even infants perform better if they are given the opportunity to compare two similar but distinct category members (Fagan, 1978; Gentner & Namy, 1999). For instance, in Oakes and Ribar’s (2010) study, 4-month-old infants showed sensitivity to categories such as ‘dogs’ and ‘cats’ when they were given an opportunity to compare the objects in pairs, rather than when instances were presented individually. Consistently, in Namy and Gentner’s (2002) study, pre-schoolers were presented abstract word categories, and they demonstrated shallower responses when the words were presented as single category members. They concluded that comparison facilitates categorization as well as conceptual development, but this occurs only when the target words have relational commonalities.

Research investigating the impacts of information presentation has dominantly concerned with learning and its implications for educational practices rather than reasoning. Evidently, organisation of information has been a longstanding focus of inductive
category learning research. Most commonly studied concept is the spacing effect, which refers to temporal delay, emphasizing memory effects in category learning. Within this strand, studies have explored the advantages of spacing repeated presentations of the same information, with learning distributed across time instead of a single massed episode (see e.g. Delaney, Verkoeijen, & Spirgel, 2010; Kornell, 2009; Proctor, 1980). Numerous studies have highlighted positive effects of this distributed information for vocabulary, memorizing facts, and word list learning (see Sobel, Cepeda, & Kapler, 2011). Other studies have considered the link between memory and previous knowledge, showing that what we know and believe influences how information is coded and remembered (see Hudson & Fivush, 1991; Ornstein et al., 1998; Sutherland et al., 2003). Other studies again have investigated as to whether larger or smaller spacing intervals increased learning and recall (Carpenter & DeLosh, 2006; Pyc & Rawson, 2009); how massed presentations increase familiarity (Dellarosa & Bourne, 1985), and whether longer delays bring memory advantages (Bjork and Allen, 1970; Cuddy and Jacoby, 1982). This parallels DeMarie-Dreblow’s (1991) general knowledge hypothesis, suggesting that increases in knowledge are associated with better memory.

Alternating presentations of categories also has positive impact on inductive learning. Kang and Pashler (2012) investigated how people can learn artist styles better. The authors found that interleaving of different artists’ paintings (presenting one item from one artist followed by an item from another artist) rather than massed presentations (presenting the same artist’s paintings in a short time period) improved inductive learning. Taylor and Rohrer (2010), using a similar design for mathematical operations in primary school students, also found that when the categories are as few as four or less, interleaving different category items results in greater benefit. This kind of presentation probably allows participants to identify the features of the exemplars and distinguish between categories (Goldstone, 1996), again indicating that contrasting comparisons are helpful. However, it also increases the temporal delay between repetitions of the same category, and requires more effortful recall over an extended period of time (Carvalho & Goldstone, 2015b). In more complex situations, rather than interleaving, blocked presentations (grouping items from the same category together) are more advantageous then (Goldstone, 1996).
For a complete conceptualization of these effects, Carvalho and Goldstone (2014) investigated whether the advantage of interleaved presentation over blocking depended on the learning situation, where category structures varied in their complexity. In order to compare ‘high between-category similarity’ (where different categories share most of their features e.g. alligators and crocodiles) and ‘low within-category similarity’ (exemplars within one category share very few features e.g. the category of ‘animal’), they used two sets of six categories, each composed of 16 exemplars of highly abstract stimuli (blobs randomly generating curvilinear segments and species of alien cells). Participants studied categories from one of these sets in an interleaved sequence and the others in blocked fashion. The authors found that three factors interacted to determine the success of inductive category learning: the structure of the category (high vs. low similarity), the presentation schedule for the categories (interleaved vs. blocked), and the temporal distribution (successive vs. simultaneous). For low similarity categories, blocked presentation resulted in better generalization, suggesting that interleaving emphasizes the distinguishing of differences between categories; whereas blocking emphasizes discovery of commonalities within the same category. Simultaneous presentation was more beneficial for later generalization of high-similarity categories, but for low-similarity categories, no difference was found between simultaneous and successive presentation.

Taken together, work on event presentations and category learning suggests that: (1) repeated experience is likely to lead to better elaborated representations of causal phenomena; (2) comparison of contrasting instances between and within types of phenomena may further facilitate elaboration and generalisation; (3) if the number of phenomena involved is small, and similarity between instances of the same phenomena is relatively high, interleaved/distributed presentation or use of simultaneous exemplars are likely to be more effective. However, these points only serve as a guide. They hardly define conditions that may promote the acquisition of better representations, leaving us with an open question as to whether these are also more supportive for causal inference.
2.4 The present study

The present study focused on the impact of information presentation on reasoning about temporally extended causal processes, and included 5 to 11-year-old children, using a unique design that included three causal phenomena (sinking, absorption, solution, inspired from physics, biology, and chemistry), in which children were presented with instances of each of these under three conditions.

In the first condition (paired-contrasting), two objects were presented at the same time for each phenomenon in turn (similar to simultaneous blocked presentation but of different category members) – e.g. a stone and a grape dropped in a tank of water – with a hypothesis that children are drawn to wonder about the invisible components of the underlying mechanism when they compare different instances of the process that unfold in different ways. This difference concerns the speed of the process, e.g. a stone sinks rapidly, grape more slowly. This presentation implies that while both objects are exposed to the same environment, this operates with variable strength depending on the density (invisible) of the sinking objects (spatial-temporal properties) and the consequent resistance of the liquid (invisible) to their fall. The comparison should suggest that at observational level the two instances ‘go together’ but with different rate; at an inferential level without one causing the other, but with some common mechanism implied which has to be determined in some fashion. From observational to inferential, the observer needs to rewind or transform the representations of the objects and their behaviour in their environment, which may demand some memory skills. However, the simultaneous contrasting presentations minimize the memory load in terms of making a decision about cause-effect connections: although the two instances of sinking are not causally related, they still share information in terms of their behaviour in water (i.e. the role of fluid density versus object density in sinking), potentially promoting better inference as well as greater representational strength.

These outcomes may be assisted or impeded by other types of presentation. To understand better, two further conditions tested the effect of paired-contrasts, each containing two instances of the different types of events, but introducing the contrasts in
different time intervals, altering the simultaneity and interleaved distribution of experiences.

*The second condition (sequential instances)* retained blocking, but employed the same materials separately in successive order (a stone first, then, after questioning about this, a grape).

*The third condition (jumbled instances)*, introduced interleaving, using the same materials with an unconnected order (a sinking stone first, then, after the same questioning, a piece of tissue paper absorbing water fast, then a spoonful of table salt dissolving quickly in water, then a grape sinking etc.).

The perceptual elements and memory load differed between the three conditions, though everything else remained the same: the paired-contrasting condition required less memory and relied more on observation; whereas the sequential and jumbled instances conditions demanded more memory but provided more space for the construction and use of representations. They also made it possible to compare children’s inferences after a single and a paired or second experience.

Children’s responses were coded at three levels, prediction from prior knowledge, description and inference, to further explore as to whether prediction or descriptive ability better assist explanation of phenomena in the paired, sequential, or jumbled design. Children’s generic verbal and nonverbal abilities were also examined in order to understand better how far individual variation in these impacted on performance under the different conditions.

The target is to ascertain which condition allows more consistent access to inference of mechanism, and to identify what processes are involved. Therefore, the three experiments emphasize causal processes rather than events, addressing two main questions below to test the hypothesis that children’s causal inference is affected by information presentation:
(1) Whether children can move directly to causal inference from immediate observation of events, or if they need to build up interim representations from distributed presentations as a step towards inference of mechanism;

(2) Whether plural instances (similar but contrasting) serve better for children to produce causal inferences than single instances. These investigations of course also required consideration of the role of prior knowledge based on past experience/observation and description in children’s performance; whether effects vary as a function of children’s age; and how far generic verbal and nonverbal ability support performance under different conditions.

The following possibilities were projected: (1) if the capacity to discern and extrapolate from paired instances is better in the sequential and/or jumbled conditions, this will suggest that having sufficient spacing to organize representations is what matters, rather than interleaving per se, as the separated instances allow children to form interim representations between direct observation and inference about phenomena they observe; (2) if there is no significant difference between the conditions, this will suggest children are able to conceive of causal processes independent of any specific form of information organisation; (3) if children’s explanations improve between seeing the first and second instances of each phenomenon in the sequential and jumbled conditions, this will indicate that witnessing contrasting exemplars is important for promoting inference, even when these experiences are temporally separated in the processes.

The study is therefore novel. No analysis is currently available on whether primary age children’s causal judgments vary depending on the presentation of causal phenomena, or whether organization of representations changes children’s causal inference, or how children encode rapid information from on-going events across time without intervening in them, and how this changes across development. The results may inform structuring of learning processes and determination of the actual source of influence where grasp of mechanisms is involved.
2.5 Method

2.5.1 Design

The study utilised an experimental design, employing three age groups spanning the English primary school age range. Five tasks were given to all children in the same fixed order within a single one-to-one session: measures of verbal and nonverbal ability, followed by the three causal tasks – sinking, absorption, solution – focused on causal reasoning, delivered under one of three conditions: (1) paired-contrasting, (2) sequential, or (3) jumbled instances. Testing sessions started with the verbal ability task to encourage children to express their thoughts as much as possible. In the second stage, nonverbal task was administered to stimulate children’s nonverbal competences along with the verbal, which are equally important to their reasoning. Causal tasks were presented in the order of sinking, absorption and solution, increasing in the administration speed respectively. Children were informed about the task order and the procedure at the beginning.

2.5.2 Participants

In total, 156 children were recruited from five primary schools, two in London, three in Oxford, covering a range of socioeconomic backgrounds [54 Year 1 (Y1), mean age = 6.5, sd = 4.5 months; 53 Year 3 (Y3), mean age = 8.5, sd = 4.9 months; and 49 Year 5 (Y5), mean age = 10.2, sd = 4.9 months]. The causal tasks piloted by working with five children (representative age range), but their results were not included in the analyses.

Only participants whose parents signed a consent form were included. The procedure and consent letter were approved by the Institute of Education Research Ethics Committee, and the head teachers and science coordinators of the five schools. Children were randomly assigned to conditions at point of testing, with 53 in Condition 1 (18 Y1, 18 Y3, 17 Y5), 50 in Condition 2 (18 Y1, 17 Y3, 15 Y5), and 53 in Condition 3 (18 Y1, 18 Y3, 17 Y5); power to detect medium effects in ANOVA=.80; in regression=.99.

Children’s home background was considered with a brief questionnaire. In total, 86 children out of 142 (60.6%) had a monolingual (only English) home environment, and 56
(39.4%) came from either bilingual or trilingual homes, indicating that the sample included a wide range of ethnic and linguistic variation. In terms of socioeconomic home environment, the sample showed a normal distribution, with 6 out 150 (4%) parents unemployed, 20 (13.3%) manual workers, 101 (67.3%) self-employed or non-manual workers, 23 (15.3%) professionals. Correspondingly, 5 out 141 (3.5%) parents had only GCSE (basic level secondary school) qualifications; 15 (10.6%) parents had A levels (higher secondary); 47 (33.3%) parents had an undergraduate degree; 54 (38.3%) parents had a postgraduate degree or professional qualification; and 20 (12.8%) parents had doctoral degrees.

2.5.3 Materials and Procedure

Test sessions took place out of class in a quiet area within schools. Each took an average of approximately 18 minutes (min = 11, max = 34). All responses were recorded manually on score sheets at the time, but children’s statements of the causal tasks were also audio-recorded for later checking.

2.5.3.1 Causal tasks

The causal tasks took place in the order sinking, absorption, and solution (with variation between conditions in the way the elements of these were presented), each involving demonstration of a contrast between two exemplars of the target process. For sinking, this was between a stone and a grape of the same size but different densities, which sank at different rates in a container of water. For absorption, it was between the rise of water from a petrie dish through strips of tissue and blotting paper the same length/width, the water rising faster through the more open structure of the former. For solution, it was between very small quantities of table and rock salt placed in warm water, the greater surface area/compactness to volume of the former leading to it dissolving more rapidly (see Figure 2).

This design controlled for the amount of information presented, but systematically manipulated the immediacy of its availability at the point of making comparisons and concluding inferences about each process.
<table>
<thead>
<tr>
<th>Tasks</th>
<th>Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinking</td>
<td>(stone, blueberry)</td>
</tr>
<tr>
<td>Absorption</td>
<td>(blotting paper, tissue paper)</td>
</tr>
<tr>
<td>Solution</td>
<td>(table salt, rock salt)</td>
</tr>
</tbody>
</table>

Figure 2. Materials used in the causal tasks

2.5.3.1.1 Condition 1 (paired-contrasting instances)

The contrasting materials in each pair were presented simultaneously, e.g. the stone and grape / the tissue and blotting paper / the table and rock salt were introduced together and children were asked to say whether they thought the same thing would happen to each; whether or not the same thing had happened; and why things had turned out differently.

For sinking, children saw a stone and a grape of similar size/colour but different densities, which dropped together at the same time and sank at different rates in a half-meter tall large transparent jar of still water. Children needed to predict outcomes ahead of witnessing simultaneous demonstration of the two instances, then to describe e.g. the rates of sinking, and to explain the outcomes, as a measure of causal inference assessing identification of basic factors (e.g. softness/hardness/heaviness of the materials), operative variables (e.g. relative weight of the materials) and mechanisms (e.g. object
density, the role of water). A typical testing session involved the following scripts:

Prediction

Experimenter: “I have got these two objects here: a stone and a grape. Do you want to hold them?” (A pause) “I am going to drop them in the water. What do you think will happen?”

“Do you think that the same thing will happen to both objects?”

Description

Experimenter: “Okay. Thank you. Please watch carefully now. I will drop the objects and you will watch.”

The experimenter drops the items together at the same time and asks: “Did you notice anything, what?”

“Did the same thing happen to both objects?”

Explanation

Experimenter: “Okay thank you. Why do you think things happened that way?

“Do you think there might be another reason for that?”

For absorption, children saw water rising from a petrie dish through strips of tissue and blotting paper of the same length/width, the water rising faster through the more open structure of the tissue. Similar to sinking protocol, children needed to predict outcomes ahead of witnessing simultaneous demonstration of the two instances, then to describe e.g. the rates of absorption, and to explain the outcomes, as a measure of causal inference assessing identification of basic factors (e.g. softness/hardness of the materials), operative variables (e.g. relative softness/hardness/porousness of the materials) and mechanisms (e.g. relative size of the holes allowing water to rise up). A typical testing session involved the following scripts:
Prediction
Experimenter: “I have got these two strips of paper here: a tissue and a blotting paper. Do you want to touch them?” (A pause) “I am going to dip them in the water. What do you think will happen?”

“Do you think that the same thing will happen to both papers?”

Description
Experimenter: “Okay. Thank you. Please watch carefully now. I will dip the papers and you will watch.” The experimenter dips the end of the papers in water together at the same time and asks: “Did you notice anything, what?”

“Did the same thing happen to both papers?”

Explanation
Experimenter: “Okay thank you. Why do you think things happened that way?”

“Do you think there might be another reason for that?”

For solution, children saw the same small quantities of table and rock salt dissolve in warm water. The small quantity of the salt was assured with two equally very small spoons. The greater surface area to volume of the table salt led to more rapid solution. Similar to sinking and absorption protocols, children needed to predict outcomes ahead of witnessing simultaneous demonstration of the two instances, then to describe e.g. the rates of absorption, and to explain the outcomes, as a measure of causal inference assessing identification of basic factors (e.g. softness/hardness of the materials), operative variables (e.g. relative softness/hardness/compactness of the materials) and mechanisms (e.g. relative size of the grains allowing water to penetrate in). A typical testing session involved the following scripts:

Prediction
Experimenter: “I have got these two kinds of salts here: table salt and rock salt. I am
going to drop a tiny piece from each in the water. What do you think will happen?”

“Do you think that the same thing will happen to both kinds of salt?”

Description
Experimenter: “Okay. Thank you. Please watch carefully now. I will drop the salts and you will watch.” The experimenter drops the samples of the salts together at the same time and asks: “Did you notice anything, what?”

“Did the same thing happen to both kinds of salt?”

Explanation
Experimenter: “Okay thank you. Why do you think things happened that way?”

“Do you think there might be another reason for that?”

2.5.3.1.2 Condition 2 (sequential instances)
The contrasting materials for each mini-experiment were presented in sequential order (e.g. the drop of the stone was followed by the drop of the grape, and children made responses about these separately). When the second item was presented, children were invited to make explicit comparison to the first as part of their predictions, descriptions and explanations.

For sinking, children saw the same materials in the sequential order. Otherwise all other details were as in the first condition. A typical testing session involved the following scripts:

Experimenter: “I have got this stone here. Do you want to hold it?”

“I am going to drop it in this water. What do you think will happen?” (Prediction)
Experimenter: “Okay. Thank you. Please watch carefully now. I will drop it in the jar and you will watch.” “Did you notice anything, what?” (Description)
Experimenter: “Why do you think things happened that way?” (Explanation)
Experimenter: “This time, I will show you this grape. Do you want to hold it?”
“I am going to drop it in the water again. What do you think will happen?” (Prediction)

Experimenter: “Now, I will drop it in the jar and you will watch.” “Did you notice anything, what?” (Description)
Experimenter: “Why do you think things happened that way?” (Explanation)
“Do you think there might be another reason for that?”

For absorption, children saw the same materials (tissue and blotting paper) in the sequential order. Otherwise all other details were as in the first condition.

Experimenter: “I have got a piece of tissue paper here. Do you want to touch it?”
“I am going to dip the end of this paper in the water. What do you think will happen?” (Prediction)
Explainer: “Please watch carefully now. I will dip the end of it and you will watch.”
“Did you notice anything, what?” (Description)
Experimenter: “Why do you think things happened that way?” (Explanation)

Experimenter: “This time, I will show you a piece of blotting paper. Do you want to touch it?”
“I am going to dip the end of the paper in the water again. What do you think will happen?” (Prediction)
Experimenter: “Now, I will dip the end and you will watch.” “Did you notice anything, what?” (Description)
Experimenter: “Okay thank you. Why do you think things happened that way?”
“Do you think there might be another reason for that?” (Explanation)

For solution, children saw the same small quantities of table and rock salt dissolve in
warm water in the sequential order. Otherwise all other details were as in the first condition. A typical testing session involved the following scripts:

Experimenter: “I have got a small amount of table salt here. Do you want to touch it?”
“I am going to drop a tiny amount of this salt in the water. What do you think will happen?” (Prediction)
Experimenter: “Please watch carefully now. I will drop a tiny amount of table salt in water and you will watch.” “Did you notice anything, what?” (Description)
Experimenter: “Why do you think things happened that way?” (Explanation)

Experimenter: “I will now show you this rock salt. Do you want to touch it?”
“I am going to drop a tiny amount in the water again. What do you think will happen?” (Prediction)
Experimenter: “Now, I will drop a tiny amount of rock salt and you will watch.” “Did you notice anything, what?” (Description)
Experimenter: “Okay thank you. Why do you think things happened that way?”
“Do you think there might be another reason for that?” (Explanation)

2.5.3.1.3 Condition 3 (jumbled instances)
The initial items of each pair of contrasting materials were presented in sequence, followed by the second item. The sequence was the stone, tissue paper, table salt, then grape, blotting paper, and rock salt. As in Condition 2, children made responses about each item separately. For three phenomena, children saw the same materials, but in jumbled order. A typical testing session involved the following procedure:

Experimenter: “I have got this stone here. Do you want to hold it?”
“I am going to drop it in this water. What do you think will happen?” (Prediction)
Experimenter: “Please watch carefully now. I will drop it in the jar and you will watch.”
“Did you notice anything, what?” (Description)
Experimenter: “Why do you think things happened that way?”

“Do you think there might be another reason for that?” (Explanation)

Experimenter: “I have got a piece of tissue paper here. Do you want to touch it?”

“I am going to dip the end of this paper in the water. What do you think will happen?” (Prediction)

Experimenter: “Please watch carefully now. I will dip the end of it and you will watch.”

“Did you notice anything, what?” (Description)

Experimenter: “Why do you think things happened that way?”

“Do you think there might be another reason for that?” (Explanation)

Experimenter: “I have got a small amount of table salt here. Do you want to touch it?”

“I am going to drop a tiny amount of this salt in the water. What do you think will happen?” (Prediction)

Experimenter: “Please watch carefully now. I will drop a tiny amount of table salt in water and you will watch.” “Did you notice anything, what?” (Description)

Experimenter: “Why do you think things happened that way?”

“Do you think there might be another reason for that?” (Explanation)

Experimenter: “This time, I will show you this grape. Do you want to hold it?”

“I am going to drop it in the water again. What do you think will happen?” (Prediction)

Experimenter: “Now, I will drop it in the jar and you will watch.” “Did you notice anything, what?” (Description)

Experimenter: “Why do you think things happened that way?”

“Do you think there might be another reason for that?” (Explanation)
Experimenter: “This time, I will show you a piece of blotting paper. Do you want to touch it?”

“I am going to dip the end of the paper in the water again. What do you think will happen?” (Prediction)

Experimenter: “Now, I will dip the end and you will watch.” “Did you notice anything, what?” (Description)

Experimenter: “Why do you think things happened that way?”

“Do you think there might be another reason for that?” (Explanation)

Experimenter: “This time, I will show you this rock salt. Do you want to touch it?”

“I am going to drop a tiny amount in the water again. What do you think will happen?” (Prediction)

Experimenter: “Now, I will drop a tiny amount of rock salt and you will watch.” “Did you notice anything, what?” (Description)

Experimenter: “Why do you think things happened that way?”

“Do you think there might be another reason for that?” (Explanation)

2.5.3.1.4 Scoring system

Each basic step in the experiments had the same three-stage structure: children inspected the materials, and then (1) predicted what would happen to the objects during the demonstration (prior knowledge); (2) watched carefully and described what they saw; (3) explained why they thought things had happened the way they had seen (inference). At each stage, they were encouraged to give as full an answer as they could.

Prediction and descriptions were scored 0-2 for accuracy of anticipating/reporting different sinking/absorption/solution rates. Inference responses were given incremental scores 0-3. Identifying only one relevant factor scored as 1 (e.g. “Both of them are heavy, that is why they sank to the bottom”; “They are papers, papers suck water”, “Salt disappear in the water”). Identifying variation with respect to these as determining rate of process scored as 2 (e.g. “The stone is heavier than the grape, it sank to the bottom”...
faster.”; “The tissue paper more softer than the other one, it gets wet easier and quicker,”; “The rock salt is bigger than the table salt, it disappears later.”). When the child described any intervening factor relevant to the causal mechanism involved, this scored as 3 (e.g. “The stone is heavier and denser than the grape, the grape is more squishy and less stuff in it. That is why stone sank to the bottom before the grape.” “The tissue paper is softer and loose that it has more space for water to climb up against gravity. The blotting paper is harder and less space inside for water to climb up. But it is not a metal or stone, it can still absorb some water.” “The rock salt is compact, and has bigger surface area. The table salt particles are smaller that water can penetrate in easier and quicker than the rock salt”).

Children’s responses were scored by two researchers independently by taking account the above identification styles. Table 1 elaborated the scoring system used by the two researchers. The independent scores were compared for the inter-rater reliability. Any difference in the independent scores was followed by further checking of the audio records, with a discussion to get a 100% agreement on the final scores.

Table 1. Scoring system for causal tasks

<table>
<thead>
<tr>
<th>Component</th>
<th>Sinking</th>
<th>Absorption</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction from prior knowledge (0-2)</td>
<td>Correct prediction for stone (i.e. sinks)=1</td>
<td>Correct prediction for tissue paper=1</td>
<td>Correct prediction for table salt=1</td>
</tr>
<tr>
<td></td>
<td>Correct prediction for difference between stone and grape (i.e. sink at different speeds)=1</td>
<td>Correct prediction for difference between tissue and blotting paper=1</td>
<td>Correct prediction for difference between table and rock salt=1</td>
</tr>
<tr>
<td>Description of observation (0-2)</td>
<td>Correct description for stone=1</td>
<td>Correct description for tissue paper=1</td>
<td>Correct description for table salt=1</td>
</tr>
<tr>
<td></td>
<td>Correct description for grape=1</td>
<td>Correct description for blotting paper=1</td>
<td>Correct description for rock salt=1</td>
</tr>
<tr>
<td>Inference/explanation (0-3)</td>
<td>No/irrelevant explanation=0</td>
<td>No/irrelevant explanation=0</td>
<td>No/irrelevant explanation=0</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Weight/size without difference between objects=1</td>
<td>Thickness/softness/texture etc. without difference between types of paper=1</td>
<td>Grain/size etc. without difference between types of salt=1</td>
<td></td>
</tr>
<tr>
<td>Weight/size with difference=2</td>
<td>Thickness/softness/texture etc. with difference=2</td>
<td>Grain/size etc. with difference=2</td>
<td></td>
</tr>
<tr>
<td>Density and mechanism=3</td>
<td>Nature of papers/holes and mechanism=3</td>
<td>Grain/size etc. with surface area and mechanism=3</td>
<td></td>
</tr>
</tbody>
</table>

Three types of measures were computed:

1. Totals across the three tasks for prediction (maximum = 6), description (maximum = 6) and inference (maximum = 9).

2. A total composite score for causal performance across these indices (alpha = .736), which could range from 0 to 21. A count was also made of the number of mechanism level responses made by children across the three conditions when drawing their concluding inferences for each contrast (maximum = 3).

3. A separate score was count for initial inference across the three tasks (maximum = 9) for children in Conditions 2 and 3. In condition 1, children saw the contrast items together. But in condition 2 and 3, they saw the pairs separately (e.g. stone and then grape, tissue paper and then blotting paper). It is possible that children’s responses would improve after seeing the second item. In order to examine whether there was any change between the first and concluding inference responses when the latter were delayed, the initial inference score was count separately.
2.5.3.2 Measures of verbal and nonverbal ability

The expressive vocabulary and block design subtests from the Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 2011) were used to provide measures of verbal and nonverbal ability. Administration and scoring for both followed standard procedures.

WASI vocabulary is a measure of expressive language, word knowledge, and verbal concept formation. There were three picture items at the beginning, where children were required to name the pictures in the stimulus book. Subsequently, children were required to give definitions for the words which the examiner read aloud. Administration and scoring followed standard procedures.

The WASI Block Design is a subtest to explore children’s nonverbal IQ. Children were shown nine red and white square blocks and a spiral booklet of cards illustrating different patterns that could be made with the blocks, and were asked to arrange the blocks to match each design in turn. The Block Design aimed to measure children’s ability to analyze and synthesize abstract stimuli within a specified time limit.

2.6 Results

2.6.1 Comparability of conditions

One-way ANOVAs showed that there were no significant differences between the three conditions in the age of participants, their vocabulary and block design scores, or their parents’ occupation and education. The length of test sessions was significantly longer for the jumbled instances condition (mean=20.06 minutes) than for paired-contrasting (mean=16.30) and sequential (mean=17.82), F(2,154)=10.800, p<.001, reflecting the greater gap necessitated by trial to trial switching in the focus of the causal task.

2.6.2 Effects of condition on causal task performance

Figure 3 plots of the means for each age group and conditions, and Table 2 shows the values for means and standard deviations, along with those for vocabulary and block
design. There was negative skew on all the measures except mechanism, reflecting a longer tail on scores in the youngest age group. Mechanism was positively skewed for the opposite reason.

Figure 3. Scores by age groups and condition on total causal performance (max=21), prediction, description (max=6), inference (max=9), and initial inference (max=9; conditions 2 and 3 only)

To bear in mind that condition 1 refers to contrasting, condition 2 sequential, and condition 3 refers to jumbled instances. Differences between conditions on the causal indices were examined using two-way ANOVAs (age group x condition). There were no effects of condition or interaction with age group on any of them, although there were consistent effects of age group: for causal total, $F=21.410$, partial eta-squared=.226; for prediction, $F=13.422$, partial eta-squared=.154; for description, $F=9.436$, partial eta-squared=.114; for inference, $F=20.429$, partial eta-squared=.217; df=2,147, p<.001 for all; $Y_1 < Y_3 = Y_5$ in each case.

In general, children performed well on description, slightly less well on prediction, but worse (in terms of the scale employed) on inference, with responses predominantly focused on differences in salient variables, and identification of mechanisms comparatively rare, though increasing as they get older. The effects of age group and the
absence of effects of condition were confirmed using the Welch robust test, which returned the same outcomes for each of the causal measures.

Table 2. Mean score with sd by age group and condition on total causal measure (max=21), prediction, description (max=6), inference (max=9), mechanism (max=3), initial inference (max=9; conditions 2 and 3 only), vocabulary (max=43), block design (max=58).

<table>
<thead>
<tr>
<th></th>
<th>Y1 (sd)</th>
<th>Y3 (sd)</th>
<th>Y5 (sd)</th>
<th>Total (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Condition 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>11.06 (3.57)</td>
<td></td>
<td></td>
<td>13.79 (3.80)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>14.50 (3.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>15.94 (3.03)</td>
<td></td>
</tr>
<tr>
<td><strong>Condition 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>12.17 (4.03)</td>
<td></td>
<td></td>
<td>14.46 (3.72)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>15.35 (3.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>16.20 (2.68)</td>
<td></td>
</tr>
<tr>
<td><strong>Condition 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>12.00 (4.04)</td>
<td></td>
<td></td>
<td>14.23 (3.90)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>15.39 (3.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>15.35 (3.02)</td>
<td></td>
</tr>
<tr>
<td>Prediction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Condition 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>3.22 (1.21)</td>
<td></td>
<td></td>
<td>4.17 (1.50)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>4.44 (1.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>4.88 (1.36)</td>
<td></td>
</tr>
<tr>
<td><strong>Condition 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>3.78 (1.21)</td>
<td></td>
<td></td>
<td>4.26 (1.21)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>4.35 (1.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>4.73 (1.16)</td>
<td></td>
</tr>
<tr>
<td><strong>Condition 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>3.28 (1.81)</td>
<td></td>
<td></td>
<td>4.06 (1.60)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>4.22 (1.48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>4.71 (1.16)</td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Condition 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>4.28 (1.49)</td>
<td></td>
<td></td>
<td>4.92 (1.22)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>5.06 (1.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>5.47 (0.62)</td>
<td></td>
</tr>
<tr>
<td><strong>Condition 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>4.44 (1.34)</td>
<td></td>
<td></td>
<td>5.06 (1.11)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>5.18 (0.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>5.67 (0.72)</td>
<td></td>
</tr>
<tr>
<td><strong>Condition 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>4.72 (1.23)</td>
<td></td>
<td></td>
<td>5.00 (1.21)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>5.17 (1.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>5.12 (1.16)</td>
<td></td>
</tr>
<tr>
<td>Inference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Condition 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>3.56 (1.82)</td>
<td></td>
<td></td>
<td>4.70 (1.89)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>5.00 (1.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>5.59 (1.73)</td>
<td></td>
</tr>
<tr>
<td><strong>Condition 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>3.94 (2.13)</td>
<td></td>
<td></td>
<td>5.14 (1.87)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>5.82 (1.47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>5.80 (1.21)</td>
<td></td>
</tr>
<tr>
<td><strong>Condition 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>4.00 (1.71)</td>
<td></td>
<td></td>
<td>5.17 (1.79)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>6.00 (1.41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>5.53 (1.66)</td>
<td></td>
</tr>
<tr>
<td>Mechanism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Condition 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>0.06 (0.24)</td>
<td></td>
<td></td>
<td>0.28 (0.66)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>0.33 (0.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>0.47 (0.94)</td>
<td></td>
</tr>
<tr>
<td><strong>Condition 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>0.11 (0.32)</td>
<td></td>
<td></td>
<td>0.30 (0.58)</td>
</tr>
<tr>
<td>Y3</td>
<td></td>
<td>0.47 (0.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y5</td>
<td></td>
<td></td>
<td>0.33 (0.49)</td>
<td></td>
</tr>
</tbody>
</table>
A three-way mixed ANOVA (time x age group x condition) on children’s initial versus concluding inferences in Conditions 2 and 3 showed that initial inferences had significantly lower scores in these, $F(1,84)=156.377$, $p<.001$, partial eta-squared=.651, indicating that children improved their inferences considerably after seeing the second relevant object. Seeing the contrasting instances therefore had a substantial effect on level of children’s inferences at all ages. There was also a main effect of age group, $F(2,84)=10.991$, $p<.001$, partial eta-squared=.207; and a modest time x age group interaction, $F(2,84)=4.094$, $p=.020$, partial eta-squared=.089, due to initial/concluding differences being slightly greater in Y3. There was no interaction with condition. Since there was no difference between conditions for concluding inference scores, it apparently made no difference whether this comparison was simultaneous, sequential or jumbled – despite the longer intervening interval and administration time in Condition 3.

Despite the lack of differences in outcome, there were indications of variation in processing across the three conditions. Controlling for age, partial correlations within each condition showed that overall, description and inference were related, as were
inference and mechanism, but description and mechanism were consistently unrelated. In Condition 1 (Table 3), prediction was uncorrelated with description and mechanism, and only modestly related to inference.

Table 3. Partial correlations between causal measures, controlling for age (significant associations in bold)

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Prediction</th>
<th>Description</th>
<th>Inference</th>
<th>Mechanism</th>
<th>Initial inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1</td>
<td>Prediction</td>
<td>.272</td>
<td>.320*</td>
<td>.121</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>.617***</td>
<td>.224</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>.653***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition 2</td>
<td>Prediction</td>
<td>.586***</td>
<td>.625***</td>
<td>.317*</td>
<td>.391*</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>.652***</td>
<td>.234</td>
<td>.363*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>.539***</td>
<td>.452**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mechanism</td>
<td>.538***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition 3</td>
<td>Prediction</td>
<td>.575***</td>
<td>.474***</td>
<td>.205</td>
<td>.104</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>.592***</td>
<td>.177</td>
<td>.331*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>.668***</td>
<td>.318*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mechanism</td>
<td>.070</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*N=43 and N=46 for correlations with initial inference, due to no audio-records being available for 6 and 7 participants respectively (*p<.05, **p<.01, ***p<.001).

In Condition 2, and to a slightly lesser extent Condition 3, these relationships were notably stronger. This suggests that having space to compare the second item with the previous demonstration might inform children’s predictions, leaving these better integrated with observation and inference overall. It also suggests that representations constructed during the task might have more influence on children’s descriptions and
inferences than those they brought into the task from prior experience. These effects were stronger in Condition 2, though, where initial inferences were also better related to the other causal indices than in Condition 3, indicating that immediate progression to a second item/demonstration led to greater integration.

These different patterns of relationship between the causal indices within each condition were consistent in all key respects when examined separately for each age group. In particular, prediction was unrelated to the other causal indices in each age group in Condition 1, but was better integrated in Conditions 2 and 3. As overall, the correlations of prediction and initial inferences to the other indices were also stronger in Condition 2 than in Condition 3. The effects of differences in the way information was presented were therefore the same regardless of age.

Applying a Bonferroni correction for multiple familywise tests, the cutoff p-value of 0.05 was divided by the total number of comparisons (.05/3) to reflect computation of the same pairwise correlations across conditions. This correction fixed the p value to 0.01 for comparisons, and the correction made little difference. For condition 1, only description and inference correlated to each other (p<.001). In condition 2, prediction, description, inference, mechanism, and initial inference were all related (p<.001); inference and mechanism; inference and initial inference, as well as mechanism and initial inference were also correlated. In the condition 3, only prediction, description and inference were related (p<.001). This suggested that across the three conditions, children’s earlier judgments had an impact on their final inferences in condition 2 (shorter time intervals), but not in condition 3 (longer time intervals).

At the same time, the general lack of relationship between description and mechanism indicates that mechanism level responses require an insight that goes beyond immediate observation as such. Table 4 elaborates this by showing that of 41 children who gave a mechanism response to one or more causal task, 37 (90%) had description scores of 5 or 6 (i.e. they accurately captured the difference between the contrasting observations for at least two out of the three causal tasks). In contrast, of 42 children who had description scores below 5 (i.e. who generally failed to do this), only 4 (9.5%) gave a mechanism
response. However, 77 children had description scores of 5 or 6 without giving mechanism responses. Good description appears to support mechanism responses, but is not sufficient in itself.

Table 4. Frequency of children making different numbers of mechanism level inferences by level of description

<table>
<thead>
<tr>
<th>Description score</th>
<th>Lower (1 to 4)</th>
<th>High (5 or 6)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mechanism responses</td>
<td>None</td>
<td>38</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>One</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Two</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Three</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>42</td>
<td>114</td>
<td>156</td>
</tr>
</tbody>
</table>

At the same time, the way in which presentation of information is organised may have an impact on this, for older children, at least: in Conditions 1 and 2, description was related to mechanism in Y3 and Y5, r=.558 to .708, suggesting that good observation feeds through to higher levels of inference as children’s overall performance improves, in line with Table 3 and 4 – but apparently only if this is based on comparison of contrasting instances within a short time interval, since no similar effect was found in Condition 3.

### 2.6.3 Relationships between causal indices and verbal/nonverbal ability

Taking overall causal score as the dependent variable, hierarchical regressions were used to examine the influence of verbal and nonverbal ability on children’s performance in each condition. Block design had a logarithmic relationship to the causal measure, whereas expressive vocabulary was linear. Age in months was entered at the first stage of the analysis, followed by log block design (the logarithmic transform) and vocabulary at the second and third stages. The change statistics therefore highlighted the differences
made by adding new predictors to the model. Results are shown in Table 5.

Table 5. Hierarchical regression analysis with overall causal measure score as dependent variable (significant predictors in bold)

<table>
<thead>
<tr>
<th>Condition 1 (contrasting)</th>
<th>Predictor</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in months</td>
<td>.510***</td>
<td>.247</td>
<td>.008</td>
<td></td>
</tr>
<tr>
<td>Block design (log)</td>
<td>.373*</td>
<td>.213</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WASI vocabulary</td>
<td></td>
<td>.488**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AdjRsquare = .397; ΔR² = .260*** for M1; .070* for M2; .102** for M3

<table>
<thead>
<tr>
<th>Condition 2 (sequential)</th>
<th>Predictor</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in months</td>
<td>.421**</td>
<td>.373*</td>
<td>.117</td>
<td></td>
</tr>
<tr>
<td>Block design (log)</td>
<td>.138</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WASI vocabulary</td>
<td></td>
<td>.436*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AdjRsquare = .230; ΔR² = .177** for M1; .017 for M2; .083* for M3

<table>
<thead>
<tr>
<th>Condition 3 (jumbled)</th>
<th>Predictor</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in months</td>
<td>.362**</td>
<td>.258</td>
<td>.129</td>
<td></td>
</tr>
<tr>
<td>Block design (log)</td>
<td>.166</td>
<td>.044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WASI vocabulary</td>
<td></td>
<td>.492**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AdjRsquare = .297; ΔR² = .131** for M1; .017 for M2; .191** for M3

*p <.05. **p <.01. ***p <.001

For Condition 1, log block design was a significant predictor when first entered, explaining 7% variance, though its beta dropped to non-significant levels when vocabulary was included, explaining 10% further variance. The drop was due to the change in the degree of freedom (df) and also overlaps in the explained variance. Note
that the df was 1 for the first model, and it was 3 for the third. In contrast, for Conditions 2 and 3 it was a substantially weaker influence, explaining 1.7% variance for both, not achieving significance even without the impact of vocabulary being taken into account. Follow up analyses showed there was no interaction between log block design and vocabulary in any of the three conditions. Although the influence of verbal ability was similar across the conditions, the implication is that simultaneous comparison of the contrasting instances and immediate inference from these draws more on nonverbal ability. If there is a delay in drawing final inferences, verbal processing is more dominant. Since the correlations between log block design and vocabulary were comparable across the three conditions (.336; .399; and .211 respectively), controlling for age, with 5-10% shared variance, this would appear to be a real effect of condition.

A similar pattern was found for mechanism responses, with nonverbal ability a more substantial predictor overall and the stronger influence in Condition 1. Initial beta for log block design=.354, final=.262, against .279 for vocabulary; while verbal ability was the stronger predictor in Condition 2, initial beta for log block design=.267, final=.154, against .370 for vocabulary, and in Condition 3, initial beta for log block design=.248, final=.166, against .329 for vocabulary.

2.7 Discussion
Children witnessed contrasting instances of three familiar phenomena in a scientific context, and needed to explain unfamiliar dimensions constituting cause-effect relationships. Of course, many children may know, for instance, that a stone sinks in water, and tissue paper absorbs water, but when it comes to answering why a stone sinks faster than a grape, this was more complicated, with the majority of children lacking understanding/explanation of phenomena in depth. Thus descriptive level responses were close to ceiling, but inferences specifically mechanism level responses were rare, though they increased with age. This may be because mechanisms require consideration of both visible characteristics, and invisible intervening factors (gravity, density), and these are not readily available from mere observation. Observables in space-time may correspond to explanatory concepts, making it possible to infer causal processes. However, this may
require a tandem between suitable levels of inferential ability and active testing such as manipulations/interventions as explained below.

2.7.1 Causal inference from observation

This study required children to observe the real world causal phenomena, and addressed the question of how well they were able to make causal inferences under these circumstances, and what helped them to do so. Across both age and condition, accurate observation and description was consistently strongly related to inference (38% shared variance overall), and, as children became older and more adept, it became better related to mechanism as well, provided they saw the contrasting instances close together.

The picture for prediction was more mixed. Where this was solely knowledge brought into the tasks, its association with inference was modest (10% shared variance) overall. Where it was based on the first element of the task, its impact was comparable to that of description. Across the three conditions then, prediction was related differently to inference and mechanism, and although the low overall correlation with mechanism (1.2% shared variance) this suggested that the mechanism understanding was independent of prediction. The prediction measure assessed different modes of thinking in three conditions – possibly more elevated observations in condition 2 and 3. Overall, this indicated that children needed to combine their predictions with other cognitive abilities that allowed them to infer obvious and unseen factors of the real phenomena.

This interpretation suggests that both prediction and description helped children to infer the phenomena up to a certain level, but specifying mechanisms required further resources. Although, this study cannot take this claim further, Waldmann and Hagmayer’s (2005) finding seem to be relevant. Derived from four experiments with adults, the authors showed that people could make predictions from observationally acquired causal model without a prior instrumental learning phase (seeing vs. doing). Other potential cues such as temporal order, instructions, causal strength, or intervention can be combined with prior knowledge about causal models and people can fill in details about parameters with the data. Note that Waldman and Hagmayer’s study was based on observation of statistical data and sampled adults. It is still unknown how children can
compensate their lack of prior knowledge with other cues when they observe natural causal phenomena. At the end, the question still remains whether children can represent causal knowledge better if they intervene in the causal processes, as shown by Sloman and Lagnado (2005), where adults performed better by ‘doing’ rather than ‘seeing’ (see also Lagnado & Sloman, 2004 for the advantage of temporal cue in intervention). A further study can investigate whether intervention enables children to access mechanism level knowledge about continuous processes better.

2.7.2 Effects of comparison and information presentation

The next question was whether information presentation played a major role. Results supported the conclusion that the availability of contrasting instances was not crucial, inconsistent with effects identified in work on category learning (e.g. Namy & Gentner, 2002), as there were no main effects or interactions with conditions. However, the results need to be interpreted carefully. For instance, correlations indicated that in the second condition, where children saw sequential instances, seeing the second item helped children to improve their inferences, though this was not the case in the third condition in which they were presented with the jumbled instances. Inference levels increased – by 2 to 3 points on average – when the contrasts were available, and although the increase in the percentage of mechanism responses (17% overall) was more modest, it highlighted that seeing the contrasts facilitated this increase in the sense that it enabled some children to exhibit more advanced levels of competence. During testing it was observable that time-wise the condition 2 and 3 required longer time to administrate. The longer testing therefore demanded longer attention span. Given that children in all conditions did well on description, the length of the condition was not a primary effect on children’s comprehension, as this can be viewed as a comprehension check.

Various effects were associated with the spacing and sequence of the contrasting instances. First, the more extended spacing effect in Conditions 2 (sequential/blocked) and 3 (jumbled/interleaved) led in all age groups to better integration of prediction than the simultaneous demonstrations used in Condition 1. Increasing space appeared to allow children to construct their representations based on the first demonstration, probably enabling them to steer their responses in the second cycle. Whether these interim
representations were more influential because of their immediacy in memory or because the prior representations were impoverished in some way is uncertain, and requires further investigation. From the results, it is clear that extended presentations allowed children (of all ages) to hold a mental image in the absence of the first stimuli (e.g. the speed of the first item) and enabled them to anticipate, compare and make inferences about the speed differences of the second item. They were also able to do so accurately for description: 73.8% of children in Conditions 2 and 3 achieved scores of 5 or 6, which means that the majority had no problem with comparing their first observation with their second. Only 9.7% of children failed to make any fully accurate description responses across the three causal tasks, two thirds of these in the youngest age group. This is consistent with the studies highlighting individual differences in the development of memory skills (see e.g. DeMarie-Dreblow, 1991; Leichtman & Ceci, 1995).

At the same time, however, these integration effects were stronger in Condition 2 than in Condition 3, indicating that immediate progression to the second demonstration was more effective. One can suggest that this is a further effect of memory: the greater delay leading to poorer recall and therefore less use of newly constructed representations to guide responses. However, the lack of accuracy difference between the condition 2 and 3 description responses makes this unlikely. It is also plausible that this is a similar effect to that reported in Carvalho and Goldstone’s (2014) category learning research: interleaving of category exemplars highlights differences between categories – not relevant in the causal tasks here – whereas blocking emphasizes discovery of commonalities within the same category, which would facilitate further application of new representations. This impact of blocking may also explain why higher levels of description were more directly associated with inference of mechanism among older children – but only in Condition 2, where blocking was employed, and in Condition 1, which used simultaneous demonstrations, also previously found to benefit discovery of commonalities. More immediate comparison of the contrasting instances seem to be helpful, in particular for younger children to deal with less visible commonalities in terms of the underlying mechanism.
Accordingly, we can think that children are able to explicitly reason about causal processes independent of any specific form of information organisation. However, moving from direct observation to inference of mechanism was still not easy for primary age. They seem to have benefited forming interim representations from sequential presentations as a step towards inference of mechanism. Resonating with this, the space to construct interim representations provided by sequential blocking and interleaving was also associated with more dominant effects of verbal ability. This can be interpreted as the space is more crucial for verbally encoding these representations. However, whether the spacing effect was also crucial for awareness of these representations is not clear, as the causal tasks contained verbal explanation. It may be possible that nonverbal forms of representation are constructed more rapidly. But this can be elaborated by a further study.

2.8 Conclusion

The use of paired-contrasting presentations would (1) save time in the demonstration of more complex scientific phenomena, (2) decrease memory load, and boredom, therefore (3) potentially eliminate some of other possible confounding factors in accessing mechanism level understanding. This is consistent with the findings of Lipsitt (1961) and MacCaslin (1954) that simultaneous presentation of two objects representing two categories reveals the impact of other factors involved in them. In the present study, use of simultaneous comparisons meant that relevant information did not need to be stored in short or long-term memory, but instead required immediate attention allocation. Although this study did not find a robust condition effect, it is plausible to suggest that paired-contrasting conditions seem to reveal seen and unseen intervening factors rapidly available to observers, within a more practical demonstration protocol.

Despite the focus is on children’s ability to think about causal processes, in the causal reasoning literature no relevant studies have been found. However, findings here supported the results of some category learning research. For instance, previous work pointed that learners did not benefit from unassisted discovery (Alfieri, et al., 2011; Kirschner et al., 2006; Klahr & Nigam, 2004; Kuhn & Dean, 2004; Singer & Pease, 1978). The results of this study can provide further answer on that why:
The patterns involved in causal inference are more complex. When we control for quality of information presentation, children’s other cognitive abilities (e.g. verbal and nonverbal) may come to the fore as further determinants.

Although children observed the same phenomena within the protocol that applied in their conditions, their mental organizations/representations produced differences between how they perceived and inferred these phenomena, where perceptual resources corresponded mostly to visual/sensory stimuli, inferential resources were intentional, and effortful. Effort may be relevant to personal choice or a competence beyond the scope of this study.

Low performers relied mostly on sensory information, they seem to be satisfied reporting the observables, instead of using this information to infer the causal phenomena in dynamic fashion.

None of the children expressed the need to see the events again, which was a significant sign that they probably felt no requirement to examine these further. Since their description responses were almost at ceiling by the age of 8, at a superficial level, there were no obvious issues with their capacity to construct representations based on a single viewing. However, children’s competence of use of these representations to make inferences about causal phenomena was substantially more variable, and constitutes a much harder step.

Verbal ability seems to be highly predictive, but not to be the only means of linking specific with prior and general knowledge. Nonverbal ability also allows connections to be made between one experience and another, using perceptual information. Both abilities seem to be needed conjointly to construct representations for inferring the mechanisms underlying scientific phenomena. Nonverbal ability of this kind has been demonstrated in studies with preverbal infants (Hayne, Boniface, & Barr, 2000; Hayne, Mac- Donald, & Barr, 1997; Rovee-Collier, 1997), but the data here suggests that growth in this ability during childhood may be of considerable significance. Nonverbal ability may promote sensitivity to immediate perceptual contrasts, with greater benefit for understanding of causal mechanisms. However, the term nonverbal is a very generic form of different types of thinking. Further studies should explore the specific forms of nonverbal thinking.
3.1 Introduction

This chapter aims to provide evidence for (i) the development of children’s reasoning about continuous causal processes, (ii) the nature and development of spatial-temporal analysis, and (iii) whether spatial-temporal cognition has a specific role in supporting children’s understanding of causal processes above and beyond spatial, statistical, verbal, and nonverbal abilities by considering age related differences in children’s use of each measure, with a hypothesis that spatial-temporal cognition is the most reliable predictor of 5-to 11-year-olds’ thinking of causal processes.

3.2 The role of spatial and spatial-temporal cognition in causal thinking

Past research has largely ignored children’s ability to conjointly manipulate spatial and temporal information. Although they have clear points of connection, spatial and temporal cognition have typically been investigated separately, with limited consideration of how they might combine. The present study postulates that most spatial qualities, such as nearness, distance, length, cannot be conceived independent of time, and vice versa. Despite foundational commonalities, children’s spatial and temporal cognition have typically been investigated separately, with limited consideration of how these might combine (see e.g. Burn & McCormack, 2009; Cheng & Mix, 2014; Frick & Newcombe, 2012; Hawes, LeFevre, Xu, & Bruce, 2015; Kushnir & Gopnik, 2007; Piaget & Inhelder, 1971, for work on spatial thinking; and Friedman 1986, 1991; Hoerl & McCormack, 2018; Levin, 1982; Lohse et al., 2015; Montangero, 1992; Piaget, 1969/2006, for work on cognition of time).

This chapter proposes that it is not possible to have a holistic view about causation and causal representations without accounting for the role of space and time conjointly. Research distinguishing space and time in causal analysis seems to follow the Newtonian
paradigm, which sees space and time as independent and fixed: while the three dimensional space is the stage which things happen, time ticks based on a clock everywhere independently in universe. On the other hand, current approaches, in particular the relativistic theory, suggests that space and time are not immutable, but they are interconnected. They cannot be disentangled, thus it is not possible to talk about space without time in experienced space, or vice versa (see e.g. Al-Khalili, 1999; Eagleman, 2008; Griffiths, 2013 for a review).

The proposal is that spatial and temporal representations are inseparable, and conjointly play a major role in understanding natural causal phenomena. The theory behind this reason is that space and time are topologically continuous; both allow computations. Neither space nor time has appearances, but observables are encoded via various qualities, such as distance, duration, velocity, direction, proximity, and metrics, enabling us to experience them within four-dimensional framework conjointly. What does that mean for causal cognition is that at perceptual level, when a cause is tied to its effect via precedence, neither spatial nor temporal qualities can be excluded from the analysis. For instance, when a cue ball makes an object ball to move, any motion occur in space by means of time, thus qualities become interchangeable and fundamentally the same facets of fabric. As soon as they stay static, one can talk about spatial or temporal configurations separately, but no causal relations involved in. Recent evidence agrees with this inseparability (e.g. see Casasanto & Boroditsky, 2008; Bonato, Zorzi, & Umiltà, 2012), and elaborates that humans do not process time and space separately, but represent time as space, space as time. For that reason, the term ‘spatial-temporal’ has been used to refer to this unification, and later analyses will show whether this unification matters across development when the concern is extracting causal information from environment. There are various forms of spatial-temporal analysis, though, beyond mere perceptual. However, the most studied form is the rapid spatial and temporal configurations between the two objects (e.g. a cue and an object ball, which required various experiments and computations to elaborate the variations in the angle/speed of the cue ball that causes object ball to move straight/diagonal). This and the next chapter will explore whether other forms of spatial-temporal analysis are possible, from immature to mature ones, that can be detected via a developmental point of view.
3.2.1 Causal thinking and the role of spatial-temporal cognition

Research relevant to the role of spatial-temporal cognition in causal thinking has focused on two main strands: causal inference and mechanism - which draws heavily on inference as studied in the tradition of Hume (1739/1978) or Kant (1783); and causal perception – which puts more emphasis on perception in the tradition of Michotte (1946/1963). The way spatial and spatial-temporal properties are utilized differs within these strands.

As discussed broadly in the first chapter, the focus on mechanism is arguably common across current psychological views on causality that otherwise differ widely, from Bayesian accounts of how mechanism-based causal structure is recovered from statistical information regarding contingency and regularity, to force-dynamic accounts such as McCloskey’s (1983). Ahn et al.’s (1995) study focuses on the mechanism information in causal attribution (see also Ahn & Bailenson, 1996; Koslowski, Okagaki, Lorenz, Umbach, 1989, for evidence adults do not solely rely on covariation in their causal judgments; see also Waldmann, 2017, for a review). Park and Sloman’s (2013) work investigated what kind of information people use to infer underlying causal structures. They found that unlike Bayesian theory implies, the screen-off rule is violated when the mechanisms are the same and people use mechanistic knowledge to infer latent structures.

The earliest study on the development of thinking about the unobservable mechanisms underpinning perceived regularities published by Piaget (1930/1960, 1971/1974). However, his tasks asked children to explain either the nature of phenomena with which they had little direct contact (e.g. steam engines), or complex operational connections and transformations (e.g. the mechanism of bicycles). He concluded that young children were pre-causal, and the requisite skills emerge late. Piaget studied the development of various cognitive abilities separately (e.g. verbal, spatial, temporal, causal mechanism), but considered the link between them. In his analyses spatial and temporal cognition accompanies the causal analysis. For instance, in his 1969/2006s book, he demonstrated that primitive/early understanding of space and time was highly dependent on duration-distance judgments. He showed children two trains travelling on parallel tracks, and the majority said that the train that travelled the longer distance took the longer time. He
concluded that children did not reliably distinguish more complex spatial-temporal characteristics such as velocity until about age nine. Later, together with Inhelder (Piaget & Inhelder, 1971), he proposed that children’s perception of space developed based on a gradual construction across three periods involving sensorimotor development from birth: ‘primitive perception of space’, ‘sensory-motor space’, and ‘representational space’. During these periods, children are pre-causal and establish their causal ideas relying on observed patterns. Correspondingly, temporal ability, such as the apprehension of duration, develops late, which is a crucial component of causal understanding. In this theory, time is related to space in particular to distance, but spatial cognition has a priority in the sense that emerging first developmentally and providing basis for temporal representations across development.

Piaget’s view has been refuted by demonstrations that pre-schoolers show concern for causal mechanism if simple tasks with low verbal demands are used. And their grasp of the mechanisms improves if the phenomenon is familiar to them from everyday life or pre-test experience (Buchanan & Sobel, 2011; Bullock et al., 1982; Bullock, 1984, 1985; Shultz, 1982; Schlottmann, 1999). Bullock (1979) for instance tested preschoolers’ understanding of causal mechanism by using a long box, with two runs sloping downward: a ball rolls across one of the windows, while a light ‘rolls’ across the other window, at the end the jack pops up. In the standard experiment demonstration (spatial-temporal contiguity between the rolling ball/light and jack popping), the majority of children and adults focused on the ball as a cause. Probably they thought that the ball had a potential to produce the action. In the unconnected experiment, the jack popped up after a pause (no spatial contiguity between ball running and jack popping) but there was no pause between jack popping and light rolling. Only adults had chosen the light as a cause, probably due to their prior experience with electrical devices generated causal mechanism. In parallel with Bullock et al.’s (1982) study, working with pre-schoolers, Das Gupta and Bryant (1989); Gelman, Bullock, and Meck (1980) also found children performed well on simpler tasks such as causal (re-)ordering or completing pictorial sequences of spatial states involving e.g. pictures of an apple, a knife and a cut apple, or as in Das Gupta’s study showing pictures of wet/broken cups and asking children what had caused the change to the initial and ending objects. In Bullock et al.’s (1982) most cited study, for instance, the authors showed preschoolers a rod that pushing a series of
blocks and making a rabbit toy to off the platform (the Fred-the-rabbit). Even three year-olds predicted that changing the length of the rod (e.g. short/long rods) to touch the blocks was change the rabbit’s state, while changing the color of the rod was not. However, they also showed that 3 year-olds’ causal understanding differed from 4/5-year-olds when spatial gap appeared in experiments. Similarly, Wilkening (1981) also found that preschool children demonstrated early implicit knowledge of time, speed, and duration when the tasks involved more practical elements with which children have direct contact in their life. Parallel to these, in Buchanan and Sobel’s (2011) study children pressed one of two buttons made a light go on; the causal button had a sticker and was connected to the light by a wire or had a battery inside. Children have brought prior knowledge of what was relevant to the mechanism to bear on inference of causal effectiveness.

Emphasizing the temporal order of events where spatial elements were intrinsic, McCormack and Hoerl (2005) used a mechanism, with pressing the blue button caused a toy car, while the red button caused a marble to appear in the window. Children each time saw only one object at a time. When the button pressing curtained, younger than 5-year-olds had difficulties to remember event order. This selectivity indicated how causal influence was transmitted from cause to effect over time, beyond what was observed. Buchanan and Sobel (2011) focused on whether children’s causal judgement affected by the change in location (e.g. changing the sticker), while McCormack and Hoerl’s (2005, 2007) study focused on the temporal order of events (chronological organization of the button pressing/doll action). The crucial outcome in Bullock et al.’s, and Buchanan and Sobel’s studies is that when 3 year-olds were shown causal mechanisms (e.g. batteries, rods), they considered the mechanism information and improved their causal inference. However, in McCormack and Hoerl’s study, children below age of 5 find difficult to remember event order, even if they knew the mechanism but were not shown with the button pressing sequence. This suggests that reasoning about temporal order may be harder than reasoning about spatial order for young children. Visual elements seem to bring memory effect in mechanism level thinking. The importance of visual element for young children’s mechanism understanding also resonates with Piaget’s (1974) view arguing that when temporal order determines causal direction, children should choose antecedent events as causes. However, young children are more constrained by the
actual sequencing of events in time when they reason about two events. Piaget highlighted that children have limited capacity to reason about reversible or reciprocal transformations of objects. For instance, for young children broken dishes are irreparable, or eaten apples are gone.

There are other studies investigating the varieties of the temporal ability in more detail (see e.g. Buehner, 2014; Buehner & May, 2002, 2003; Campbell, 2006), but when the focus is the key developmental achievements regarding the maturation of temporal ability and thinking about causal processes, few studies inform us about other components influencing these. For instance, Hoerl and McCormack’s (2018) and McCormack and Hoerl’s (2005) studies highlight that there is a shift from relying on temporal updating to being able to reason about time. And the emergence of thought about other non-present time points (e.g. past, future) is necessary for the maturation. As discussed above, Piaget’s argument was that it was the lack of understanding of mechanism that constrained children to reason about temporal order. Of course the nature of the tasks are different, more child friendly in these recent studies and they underline that young children seem to be better in temporal updating than temporal reasoning. Children below 5 years seem to find it difficult to reason about time, in particular future situations (see e.g. Burns et al., 2018; Hoerl and McCormack, 2018; Friedman, 2003; Martin-Ordas, 2017); they start understanding time in an event-independent way, as an emerging ability to comprehend of time in an abstract fashion (see also Blakey et al., 2018; Friedman, 1982, 1989, 2003; McColgan & McCormack, 2008; McCormack, 2015; McCormack et al., 2018; Rattat & Droit-Volet, 2007; Tecwyn, Thorpe, & Chappell, 2014). Taking together, these studies makes us think that understanding causal mechanism might develop at different ages due to children’s uneven profiles in utilizing spatial and temporal properties during early childhood. However, we expect this profile to be much homogenously developed over primary years.

All these studies signal a potentially important connection between the development of spatial-temporal and causal analysis, but a joint analysis has received little attention. Moreover, in the early investigations, spatial-temporal analysis was not separated from
causal analysis, leaving it unclear as to whether it constitutes a distinct ability that
children learn to deploy to support causal thinking. Much other research has considered
how scientific thinking develops over the school years, but this tends to concern
metacognitive processes and strategies for evidence generation/evaluation (e.g. Klahr &
Nigam, 2004; Kuhn, 2011; Kuhn, Iordanou, Pease, & Wirkala, 2008), not causal thinking
or the abilities that underpin it.

In work on causal perception, Michotte’s (1946/1963) pioneering studies on perceptual
causality with adults documented that causal relations did not only need to be inferred
from regularities as Hume (1739/1978) suggested. Using the example of billiard balls,
one causing another to move, he found that for the vast majority of participants when a
cue ball (A) hit another ball (B), the motion of B was not perceived as its own, but
instead was perceived as a simple continuation of A’s motion, concluding that causal
perception relied highly on spatial-temporal elements, such as velocity, temporal
precedence, and directionality.

Following Michotte’s paradigm, studies demonstrated that perceptual causality appears
from infancy (Leslie, 1984; Oakes & Cohen, 1990) undergoes development with different
rates during childhood (e.g. Schlottmann, Allen, & Linderoth, 2002). Perception of object
interactions can be direct, immediate, without the assistance of prior experience,
language, or causal learning. Even small manipulations of latencies, velocities, or
direction of stimuli in the launching effect disrupt the perception of causality, indicating
that this is stimulus-driven, rather than reflecting more general inference (see e.g. Scholl
& Tremoulet, 2000). In this strand of work perceptual processes do not necessarily
require mechanism level thinking, but maturation seems to mediate inference of the
physical structure of the world.

In both strands of work (perceptual and inferential), however, causality is
characteristically limited to distinct, clearly segmented events (A causes B to happen).
This schema does not apply to continuous causal processes (e.g. the earth travelling
around the sun), where there is no contiguity between distinct events. Natural phenomena
that children encounter in school science commonly involve such temporally extended
continuous processes, as when objects sink, dissolve or soak up water. It is however unknown as to whether primary school age children find making inferences about natural mechanisms more difficult than about event-based mechanisms. As hypothesized, spatial-temporal analysis may bridge between mere observation of continuous processes and their causal analysis, enabling children to (i) extract key dimensions of information from states that change over time, effectively segmenting continuous processes into distinct steps; (ii) conceive of the sequence of dynamic transformations that underlie such observed change, and (iii) projecting these transformations onto past, present, and future experiences in order to anticipate and explain the underlying mechanisms that produce them.

To test this hypothesis, two tasks were employed to assess whether these differentially predict children’s causal thinking about these. In this paradigm, spatial-temporal analysis is distinct from causal analysis. The first task employed here, adapted from Piaget (1969/2006) ‘flow of liquid’, required only analysis of how spatial configurations changes over time. In this task children had to observe and recall segmented stages of how liquid flowed from an upper to a lower flask, marking on paper the levels in both flasks at each stage. The flask images were then scrambled and children had to reconstruct the spatial configurations in the correct temporal order. They therefore needed to consider the sequences of dynamic changes in the flask system, principally when the liquid level in the top flask decreased, it increased in the bottom flask. The other types of spatial-temporal task ‘speed’, developed by the researchers, involved various objects (bunnies) moving at different speeds from various starting points towards a common goal. In this task, children saw the beginning of the motion and had to extrapolate its end, i.e. a future spatial configuration.

3.2.2 Causal thinking and the role of spatial ability

Although there is good evidence for a link between spatial thinking and chemistry (e.g. Stull, Hegarty, Dixon, & Stieff, 2012); physics (e.g. Kozhevnikov, Motes, & Hegarty, 2007); and success in STEM – science, technology, engineering, mathematics – fields (Cheng & Mix, 2014; Newcombe, 2010, Stieff & Uttal, 2015) and recent years have seen
growing interest in spatial thinking (see Uttal et al., 2013, for a general review), its role in causal thinking has not received interest.

The varieties of spatial ability have been explored in much detail. Earlier studies found that spatial cognition was not a unitary form of thinking; instead there were several spatial abilities, such as image generation, storage, retrieval, and transformation (Carroll, 1993; Guilford, 1967; Lohman, 1996; Vernon, 1950). Further evidence demonstrated that some perceptual representations were iconic, while some of them were image-like (see e.g. Fodor, 2008; Kosslyn 1980, 1995). The former are semantically organised (i.e. decomposable sentences), while the latter are based on object perceptions, which have pictorial content and are spatially characterized (Heck, 2007). Recently, Newcombe and Shipley’s (2015; see also Uttal et al., 2013) theoretical model popularly distinguished between intrinsic/extrinsic and static/dynamic spatial abilities, as a basis for a classification of spatial tasks. Intrinsic skills involve the processing of spatial relations within objects with or without dynamic alterations in these, as in mental rotation and paper folding tasks. Extrinsic skills address spatial relations between objects, again without or with transformations of these relations. Developmentally, studies indicate that intrinsic [e.g. mental rotation as an intrinsic static (Estes, 1998); mental folding as an intrinsic dynamic ability (Harris, Newcombe, & Hirsh-Pasek, 2013)] and extrinsic abilities [e.g. spatial perspective taking (Hegarty & Waller, 2004)] develop at different rate. For example, while intrinsic abilities emerge around the age of four, extrinsic dynamic abilities are detectible later in the development. Around the age of 9/10 children show a dissociation in their abilities where intrinsic spatial ability mature into an adult form (Vander Heyden et al., 2016). However, in all of these the focus is on spatial configurations; even in dynamic tasks, the temporal element is lacking; it is unknown as to whether these potentially important abilities play a role in cognition of causal processes.

The link between high spatial ability and success in STEM careers is the most robust established relationship, but refines it to a focus on a specific type of ability. Evidence supports the notion of a distinction between types of spatial ability, such as ‘schematic spatial’ (encoding spatial relations) and ‘visual spatial’ imagery (encoding visual
appearance), and of individual variation in the use of these with consequent impact on performance. It is commonly accepted that spatial ability is highly related to schematic but not pictorial representations (e.g. Poltrock & Agnoli, 1986). For instance, while the use of schematic representations is associated with mathematics achievement, the use of pictorial representations is negatively correlated with it (Hegarty & Kozhevnikov, 1999; Khooshabeh & Hegarty, 2010).

In Shepard and Metzler’s (1971) mental rotation observers should identify a target shape among foils in rotated positions in a schematic fashion. Here the spatial configuration of the system is static, and observers find the correspondence by mentally simulating its rigid motion. The paper-folding task (Eliot & Smith, 1983) is a more complex variant. A folded piece of paper is pierced, and observers need to identify the correct configuration of the holes in the unfolded paper. Again the spatial configuration of holes in the paper does not intrinsically change, though the observer perspective does.

Such spatial tasks concern fixed systems, while spatial-temporal tasks concern systems with spatial changes over time in the system itself, involving information about distance, duration, speed, and sequence of states. In the literature these are often described as vital cues for causal analysis. However, it is unknown whether any forms of spatial alongside spatial-temporal ability have a role in causal thinking, or whether they share their variances in predicting reasoning about real world phenomena. To test this, the present study included the variants of forms of schematic spatial thinking, with most robust forms of spatial tasks. The aim is to explore further the nature of the relationships between them in predicting children’s thinking about continuous causal processes.

In doing so, two spatial tasks were included: the mental rotation and paper folding tasks. In the mental rotation task (adapted by Broadbent, Farran, Tolmie, 2014, from Shepard and Metzler, 1971), children were asked to identify a target shape among examples in rotated positions. Here the spatial configuration within the system was static, and observers had to find the correspondence by mentally simulating its rigid motion. In the paper-folding task (Eliot & Smith, 1983), children had to identify the configuration produced as a consequence of specific folds being made in a piece of paper. Although
this task contains a dynamic aspect, the spatial characteristics of the paper are not intrinsically modified. Both of these spatial tasks concern fixed systems (unlike spatial-temporal tasks that concern systems with spatial changes over time in the system itself).

3.3 Causal thinking and the role of statistical inference

Hume (1739/1978) argued that we can only know about causality from the ‘constant conjunction’ of potential causes and effects: it cannot be perceived directly, but must be inferred from the statistical co-occurrences of events. Since then, multiple schools of thought have put some form of statistical analysis of repeated experience at the core of causal thinking, ranging from the causal attribution literature in social psychology (Kelley, 1967, 1973) to work on associative causal learning inspired by animal studies (Shanks & Dickinson, 1988). However, although causes covary with their effects, there are many well-known pitfalls in inferring causation from correlation. Current models of causal reasoning therefore integrate Humean analysis from statistical information with more Kantian reasoning (transcendental conditions of experience), focused on the mechanisms that connect cause to effect (see Waldmann, 2017, for a review).

The vast amount of research has considered how causal thinking makes use of statistical regularities and knowledge about mechanisms has almost all addressed how we link distinct, separate perceptual events into a causal sequence, for instance, when pushing a button is followed by a light coming on, as discussed in the previous chapters. We do not know whether the ability to analyse covarying events or understanding of probability also assists causal thinking about continuous causal processes when they need to be registered from a single phenomenon. Statistical thinking may be important not only for extracting causal relationships from complex patterns of ‘data’, but also for considering single instances, datum, and dynamic causal processes together with spatial-temporal analysis. Given that sometimes children may need to infer causal relations from single instances (rather than intensive data) through their spatial-temporal properties, it seems to be important to understand whether they actually need to employ statistical thinking to reason about causal processes within single cases [e.g. Some may call this as single-trial learning (see Lagnado, 2011), others name it as causal induction from impoverished data (Griffiths & Tenenbaum, 2003, 2005, 2009)].
Probability (definiteness) and covariation judgments conceptually differ in their relation to causality. They will therefore be discussed separately in more depth to aid the question of whether statistical sensitivity helps children determine the strength of an effect and highlight the potential role of unseen/uncertain mechanisms.

3.3.1 How can probabilistic thinking play a role in causal processes?

Probability is a tool to deal with uncertainty, an estimate of the likelihood of an event that may or may not occur (mud suggests rain). The importance of this skill when applied to causal processes is that awareness of uncertainty highlights the existence of invisible/unobserved measurements/factors, which allow exceptions (not all mud suggests rain, but sometimes flooding). Understanding of probability may in general therefore lead to greater sensitivity to the possible role of unseen aspects of causation, which may have particular value in thinking about underlying mechanisms.

Interest in the role of probability has, of course, already led to psychological investigation, including Bayesian causal learning approaches, linking probability understanding and causal inference, both with children (e.g. Gopnik et al., 2004; Gopnik & Tenenbaum, 2007; Kushnir & Gopnik, 2005, 2007; Wu et al., 2011), and with adults (e.g. Cheng & Holyoak, 1985; Griffiths & Tenenbaum, 2009; Tenenbaum et al., 2011; see also Hagmayer et al., 2007 for the ways causal Bayes nets can support causal inferences through intervention; and Cheng, 1993; Cheng et al., 1996; Waldmann & Holyoak, 1992 for data-driven processes that generate causal knowledge; see Buchsbaum et al., 2015, for Bayesian analysis of statistical regularities in continuous action sequences). Research has also compared people’s ability to use a variety of cues, such as statistical and temporal information. For instance, Lagnado and Sloman (2004, 2006) investigated whether intervention or observation improved the efficiency of learning causal structure when using probabilistic data. They found that learners were more successful in identifying the causal model when they intervened on the system rather than restricted to observing different states of the causal network. However, this was not straightforward, as their findings also showed that the capacity to infer causal structure was not only determined by the interventional data, but also learning through
interventions was more successful when the temporal cues accompanied. Although learning is not within the focus of this thesis, this finding highlights that the temporal cue is a more stable indicator when one is intervening on a system rather than passively observing it due to the fact that the intervener experiences directly that his/her action must precede the effect.

Comparing children’s and adults’ abilities, McCormack et al. (2015) found that in particular young children (up to 7 years of age) were not able to use intervention information to judge causal chain structures, instead they preferred temporal information to distinguish between causal structures. This is in consistent with the inherent bias account, suggesting that people generally prefer temporal cues over statistical information. In this study children were asked to make causal structure judgments in a simplified learning task: in all the experiments children needed to decide whether a three-variable system was either a common cause (B\(\leftrightarrow\)A\(\rightarrow\)C) or a causal chain (A\(\rightarrow\)B\(\rightarrow\)C or A\(\rightarrow\)C\(\rightarrow\)B). This study showed that the use of probabilistic information was computationally challenging for young children and demand on memory skills (see also McCormack et al., 2016). This finding resonates with White’s (2014) argument suggesting that use of statistical information for causal inference develops late.

Piaget and Inhelder (1975) also considered probabilistic thinking as a formal operational achievement, where older children could quantify the relative proportions of target and non-target events in order to calculate probabilities. However, contrary, in Anderson and Schlottmann’s (1991); Acredolo et al.’s (1989); and Schlottmann’s (2001) studies it has been clear that children from age 4 are sensitive to probability, and can make judgments that conform structurally to the normative probability ratio. From Sobel and Kirkham’s (2006, 2007) Bayesian perspective, even infants are sensitive to subtle statistical concepts, such as the conditionalized probability relations that help adults separate causal from spurious relations. In this strand of work, studies report that even infants seem to be sensitive to probabilities when observing random, non-intentional data (e.g. altering samples of various colour balls in different boxes) (e.g. Denison & Xu, 2014). However, these recent studies do not conclude whether young children can use statistical information to grasp of causal structures, in particular of mechanisms (how things work).
This leaves us with a limited knowledge about how non-Humean view corresponds to probabilistic thinking.

The debate on ‘whether mechanistic and probabilistic analyses of causation are contrast in nature’ traces back its roots to philosophical discussions. In most cases causal mechanisms can be understood from the probabilistic implications and covariation when they rely on common cause structures (i.e. gravity keeps the moon in orbit around the Earth and causes the ocean tides). However, causal mechanisms imply generative connections, while statistical information is computational and correlational, but on a priori grounds, statistical reasoning – any analysis based on information about the frequency of occurrence or co-occurrence of events, objects or features, extracted from recurrent experience – may be an important component of causal thinking and a cue to causal mechanism.

What cognitive studies tell us about the distinction between the mechanism and probabilities then? Ahn et al. (1995) propose that casual explanations likely to produce mechanism, and given a choice, adults seek information about mechanisms rather than about statistics when making causal judgments. Adults recognize that statistical information needs to fit with the mechanism, because it is the latter that generates statistical structure, and that distractor events not generated by the mechanism may interfere with observation. This view also underlines that probability/covariation information is not a necessary condition for the grasp of causal relations (see also Ahn & Bailenson, 1996). Ahn et al.’s view contributes to the literature in clear ways. However, this approach cuts off the probabilistic implications and stipulates greater prior knowledge in causal thinking. On the other hand, probabilistic inference may be an important factor for inferring unseen/unmeasured aspects of dynamic causal processes, as discussed earlier, in particular when prior knowledge is not available. Glymour and Cheng (1998) argue about that assumption and propose that mechanism is not against the probabilistic approach. For instance, when the likelihood of moon orbit given ocean tides is greater than the likelihood of moon orbit given not-ocean tides, an intervention on ocean tides would have no effect on the likelihood of moon orbit, but the mechanism and probabilities are still there. In many situations, the underlying mechanism is unknown,
and in these, statistical information may provide some elementary clues for causal relationships. Developmental studies with mechanism view of causation can be classified as being parallel to this view (e.g. Bullock et al., 1982; Gelman, Bullock, Meck, 1980; Keil, 1979), as they follow neither Humean nor Piagetian perspective. However, none of them concerned about the distinction between probability and mechanism view.

Considering the two approaches, Humean regularity and Kantian generative mechanism, Schultz (1982) worked with 3- to 13-year-olds. He reported five experiments, where i.e. sound, wind, light transmissions were presented children in different procedures to assess the essential meaning of causation. Children received problems on each of these apparatuses: transmission from source, temporal contiguity versus generative transmission, spatial contiguity versus generative transmission, and covariation. Similar to Ahn et al.’s findings, he found that children consistently prefer generative rather than covariation information when they see a conflict between them. For instance, children’s justifications were mostly based on mechanism, rather than covariation. Most 3-year-olds’ verbal abilities were poorer at the generative aspects of the problem. However, contrary to Ahn et al.’s proposal, Schultz’s results showed that the tendency to analyze causal mechanism is not restricted to prior knowledge; children’s familiarity to the objects was not a strong indicator of mechanism level thinking (see also Koslowski, Okagaki, Lorenz, Umbach’s 1989 study that samples college students). Taken together, the literature suggests that although probabilistic/covariation information is not preferred when they are pitted against mechanism, probabilistic thinking may still be important. The reason is that children’s ability to judge probability may not just index computational ability, but also sensitivity to definiteness of outcomes in the world.

This chapter aims to contribute to these discussions in three ways: first, children’s sensitivity to probability and also their computational abilities will be observed. Second, further evidence will be provided for whether the development of probability understanding plays any role in making inferences about natural causal phenomena involving extended continuous processes. Third, competences measured by spatial-temporal and statistical tasks will be compared in the same model to reveal whether they are related or predictive when children observe a single causal phenomenon. Participants
will not intervene in the processes, but they will be encouraged to think about the causal mechanisms from the observation of three single natural phenomena. The tasks – probability, covariation, spatial-temporal- will be presented independently to elicit whether children’s computational ability or sensitivity to probability mattered for the cognition of causal processes.

The tasks in which children exhibit these abilities typically involve non-causal models, displaying all outcome possibilities simultaneously, to minimize memory requirements (i.e. in the marbles task the child sees a plate with 7 red winner marbles and 3 blue loser marbles, and judges how easy it is to win in a blind draw). In tasks where probabilities are experienced sequentially (e.g. the child draws a number of times from a population with initially unknown proportion of winner and loser marbles), children do not do so well when predicting the next outcome, as has long been known from work on probability learning (Brainerd, 1981) and child variants of the Iowa Gambling Task (Huizenga, Crone & Jansen, 2007). Children’s difficulties in sequential tasks may reflect memory capacity and other processing limitations, though, and in any case indicate problems with cumulative estimation rather than basic grasp of probability.

These two types of task address different aspects of understanding, as discussed in Schlottmann and Wilkening’s (2012) review. Classical probability derives appreciation of uncertainty and likelihood from rational analysis that multiple outcomes are possible in a given situation and from enumeration of these outcomes, prior to experiencing instances of the outcomes themselves. Probability tasks laying out all outcome possibilities simultaneously for children (e.g. showing them all the marbles on a plate, as in one of the tasks employed here) provide opportunity for such analysis. Children typically do well on these.

The alternative frequentist view derives probability from a distribution of variable outcomes over time, which ostensibly requires greater attention to the detail of that distribution. Sequential probability tasks are modelled on this, conforming to the way in which probability is often encountered in everyday life, where we may not have an a priori idea of the likelihood of an outcome, or indeed even of the fact that the outcome is
variable, until we begin to experience the situation. Even though children do not do so well on these tasks, due to higher processing demands, these skills still link to probability understanding (Bayless & Schlottmann, 2010).

Probability understanding per se comes prior to ability to calculate probabilities, which is largely established in early years, though children’s understanding of how to quantify it may be restricted to simple relations like ‘more’ or ‘larger’. Bryant and Nunes (2012) showed that more refined proportional reasoning is highly trainable regardless of children’s initial ability, and that training is effective during the elementary years, suggesting that it too is within children’s competence in this age range.

In this study, a sequential as well as a classical (proportional) probability task was employed to minimize the processing requirements of the former. An estimate of probability involves consideration of the frequency of one event A relative to other possible outcomes (e.g. ‘it will rain’; ‘of rolling a 6’), with \( p(A) = \frac{A}{A + \text{not}A} \), so this concerns one-dimensional data, while an estimate of covariation concerns two dimensional data. Whether variation in one dimension (climate change; what is inside a box) links to variation in another (human industrial activity; the colour of the box). Probability therefore may tap into a computationally simpler form of statistical reasoning than covariation.

A task with lower computational demand was needed also appropriate for the age range: the third “randomness” task was added in the battery. The randomness task addressing the related and even more basic insight, that sometimes outcomes are determined and predictable, while in other situations they are unpredictable or potentially random (Bryant & Nunes, 2012; Reyna & Brainerd, 1994). Children make this distinction from ages 4 or 5, as shown by Kuzmak & Gelman (1986), who presented children two devices, one deterministic (marbles lined up in a clear tube, with the first coming out on each trial), one a lottery device (a cage full of spinning marbles). Children understood that in the first device each outcome is known, but in the second it is not. Study 1 here employed a similar task, the distribution of target cards in shuffled and unshuffled decks, with the anticipation that this would be sensitive even to the youngest children’s abilities.
All together, three probability tasks were included, involving different levels of processing complexity, and requiring different levels of understanding. These tasks may elicit variation in performance at different ages, and clarify which task might be related to which aspect of thinking about continuous causal processes, such as relative ‘definiteness’ of effect (e.g. stones are very likely to sink, berries and grapes are less likely to) or, as noted earlier, unseen aspects of causation (i.e. some other factor affects the relative probability of sinking).

3.3.2 How can covariation information play a role in causal processes?

It may be more demanding to grasp covariation than probability, because children must track variation in not just one, but two variables, and recognize whether this indicates a link between them. Detection of such links would clearly be helpful in identifying potentially causal variables. Indeed, virtually prior studies with children use examples of causal covariations. Developmental work demonstrates that children can use covariation to infer causal relationships from a young age (see e.g. early studies by Shultz & Mendelson, 1975; Shultz, 1982; and more recent work by Schulz et al., 2008). By school age, children are clearly able to extract event regularities to make inferences about causal relations between two events. In Schulz, Gopnik, and Glymour’s (2007) study, preschoolers were shown pairs of gears (B and C) operating with a causal chain and a common cause structure on the basis of observing interventions between them. This study argued that children as early as four-year-olds discriminated between causal chain and common cause structures. Following Schulz et al.’s (2007) study, Sobel and Sommerville (2009) also found that young children are able to discriminate between causal chain and common cause structure. Like adults, when there are conflicting perceptual or mechanism cues, children prefer the latter to statistical information (Bullock et al., 1982; Shultz, 1982), as discussed earlier, though this may in part reflect limited processing resources and the demands of integrating information over multiple events.

The interest is typically on whether children grasp the implications of covariation information about distinct events for causation. These studies mostly compare the simple case of two potential causes, one regular and one irregular covariate of the effect. To
reduce processing demands, only minimal information is given, on whether a cause always co-occurs with the effect (AB cases) or whether in some instances a cause occurs without the effect (AnotB cases). If both frequencies are considered, one can derive the probability of the effect, given the cause. This, however, is only part of true covariation assessment, which also requires consideration of the base rate, the extent to which the effect occurs in the absence of the cause (i.e. notAB versus notAnotB cases, in terms of a 2x2 contingency table, as shown in Table 6).

Table 6. A typical 2x2 contingency table, with cause and effect as the two variables

<table>
<thead>
<tr>
<th></th>
<th>Blossomed</th>
<th>Dead</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant received fertilizer</td>
<td>AB</td>
<td>AnotB</td>
</tr>
<tr>
<td>Plant didn’t receive fertilizer</td>
<td>notAB</td>
<td>notAnotB</td>
</tr>
</tbody>
</table>

The literature focused on covariation (or contingency) judgment therefore considers how humans utilize information from all four cells. A well-established approach is based on the delta p statistic, which is the difference between the two probabilities discussed above (Jenkins & Ward, 1965; Dennis & Ahn, 2001; Marsh & Ahn, 2009). Adult covariation judgment is often studied by providing numerical summaries of the instances in explicit contingency tables, though the instances can, of course, also be presented sequentially, as in the real world, which adds memory demands. To avoid this, and also lower the numerical requirements of such tasks, pictorial formats are typically used with children (see e.g. Shaklee & Mims, 1981). Note that, as in Table 6, these types of studies still illustrate covariation information in causal contexts, to attempt to make complex structured data patterns intuitive and meaningful for children.

Even so, however, children commonly fail to use the delta p strategy appropriately, but instead employ simpler strategies that do not consider all four cells of the table or do not weight them evenly. Using this approach Shaklee and Mims (1981) demonstrated four strategies used by children across development, hierarchically increasing from the least to the most sophisticated: judgment of the frequency with which the target events co-occur (AB); comparison of the number of times target events do and do not co-occur (AB)
versus AnotB); comparing frequencies of events confirming and disconfirming the relationship (AB plus notAnotB versus AnotB plus notAB); and optimal assessment of the conditional probability of events (delta p).

These patterns suggest a shift from less to more accurate use of covariation data, where frequency judgment based simply on positive co-occurrence emerges early, while the conditional probability strategy does not appear until the tenth grade. Consistent with this, Shaklee and Paszek (1985) found that, in elementary school, children were most likely to make judgments about covariation by comparing frequencies of the target event, and use of the more advanced strategies identified by Shaklee and Mims was rare even in fourth grade. Ferguson et al. (1984) found that older children’s impressions were influenced more by fuller covariation information rather than frequency information, and the use of frequency information decreased with age while covariation information increased. However, even adults still have difficulty with the delta p (or related complex) covariation rules (e.g., Hattori & Oaksford, 2007).

Primary age children’s apparent tendency to focus on frequency over covariation may reflect their difficulties of understanding, but it may also be influenced to some extent by the tasks used. When computational demands (e.g. ratios, percentages) are minimized, even young children appreciate the difference between variables that co-vary perfectly with an effect or are unrelated to it (see Schulz et al.’s, 2008, experimental design with four conditions, where children observed a block hitting another block causing it to emit either a train or siren noise). Assessment of imperfect correlation poses more problems, though this is affected too by the way information is presented. For instance, in simple symmetrical tasks (asking whether green or red chewing gum causes bad teeth as illustrated over ten pictures), even four year olds could evaluate patterns of covariation (Koerber et al., 2005). Similarly, Saffran et al. (2016) found that symmetry of variables in contingency tables (present/present tables that address the outcomes of different levels of a factor are symmetrical, present/absent tables that deal with its presence versus absence are asymmetrical) had effects on data interpretation in both children and adults, both found symmetrical easier.
While in supportive circumstances even young children may be able to assess covariation from all four cells of the contingency table, if tasks have any complexity elementary school aged children fall back on strategies considering only the relative frequency of positive co-occurrence (AB versus A\(\neg\)B cells). For the present study, a more child friendly non-causal covariation task was devised, to be able to assess whether individual differences in covariation assessment predict children’s causal thinking. The link between the variables relied on the statistical information, but without support by potential knowledge of a causal relation. In all other respects, the task was as simple as possible, using a pictorial approach, consistent with the literature and with the other tasks in the battery, to investigate children’s assessment of simple covariation patterns (see Figure 9 for the example). Children had to evaluate whether a particular surround shape (e.g., circle or square) “goes with” particular picture inside (e.g. ice-ream or basketball). The focus of this task was on whether children could extract this relation in a number of problems differing in the relative frequency of co-occurrence.

3.4 Method

3.4.1 Design

The study combined a cross-sectional individual differences design, employing three groups spanning the English primary school age range. Tasks were given to children in fixed order within a single one-to-one session: measures of verbal and nonverbal ability; three causal thinking tasks; two spatial-temporal tasks; two spatial tasks, three probability (randomness, classical, and frequentist probability) and one covariation tasks.

The sample included 107 children aged 5 to 11 years; and examined whether the previously little-researched capacity for spatial-temporal analysis facilitates reasoning about continuous processes – especially with regard to inferring underlying causal mechanisms – over and above spatial, statistical, verbal, and nonverbal ability. Children were tested individually with the same causal tasks (sinking, absorption and dissolving), on a three-stage methodology (prediction, description, explanation). Further tasks assessed spatial-temporal (flow of liquid, extrapolation of relative speed,), spatial (mental rotation task, and paper folding), statistical (randomness, cards, marbles, covariation),...
verbal (expressive vocabulary), and nonverbal (block design) ability. Age dependent patterns were detected for both causal and predictor tasks.

3.4.2 Participants

In total 120 children were recruited, with parental consent, from three primary schools in London and Oxford. Subsequently 13 of them were excluded for low attention span or unwillingness to continue; one child was excluded following parental request, leaving 107 individuals for analysis [35 from Year 1 (Y1), mean age = 6 years, 1 month, sd = 4.4 months; 33 from Year 3 (Y3), mean age = 8 years, 4 months, sd = 5.9 months; 39 from Year 5 (Y5), mean age = 10 years, 3 months, sd = 5.9 months].

Of 103 participants whose parents responded, 59 (55.1%) were from monolingual (English) environments, and 44 (41.1%) from bilingual/trilingual homes, confirming the sample encompassed wide ethnic/linguistic variation. The sample was skewed towards the upper ranges of SES, with 8 (7.5%) parents being manual workers, 17 (15.9%) self-employed/non-manual workers, and 78 (72.9%) professionals. One parent (0.9%) had only GCSE (basic secondary school) qualifications; 6 (5.6%) had A levels (higher secondary); 39 (36.4%) had undergraduate degrees; 33 (30.8%) had postgraduate degrees or professional qualifications; and 24 (23.3%) had doctoral degrees.

3.4.3 Materials and procedure

Tasks were given to children in fixed order, as below, within a single one-to-one session, in a quiet location within school. Sessions lasted on average 37 minutes (min = 18, max = 45). Responses were recorded manually on score sheets, but children’s replies during the causal tasks were also audio-recorded for later checking.

3.4.3.1 Measures of verbal and nonverbal abilities

Testing started with the expressive vocabulary and block design subtests from the Wechsler Abbreviated Scale of Intelligence -WASI (Wechsler, 2011), which were used to provide measures of verbal and nonverbal ability. The reason to start with these tests was that children were encouraged to talk during the verbal task administration, and then
focused on nonverbal dimensions that warmed them up towards the causal tasks. Administration and scoring followed standard procedures (cf. Chapter 2).

3.4.3.2 Causal tasks
The verbal and nonverbal tasks were followed by the three causal tasks each highlighted a direct contrast between two instances of the target phenomenon, presented simultaneously.

For *sinking*, children saw a stone and a blueberry of similar size but different densities, which sank at different rates in a half-meter tall large transparent jar of still water. Children needed to predict outcomes ahead of witnessing simultaneous demonstration of the two instances, then to describe e.g. the rates of sinking, and to explain the outcomes, as a measure of causal inference assessing identification of basic factors (e.g. softness/hardness/heaviness of the materials), operative variables (e.g. relative weight of the materials) and mechanisms (e.g. object density, the role of water). A typical testing session involved the following scripts:

Experimenter: “I have got these two objects here: a stone and a blueberry. Do you want to hold them?”

“I am going to drop them in the water. What do you think will happen?”

“Do you think that the same thing will happen to both objects?” (prediction)

Experimenter: “Okay. Thank you. Please watch carefully now. I will drop the objects and you will watch.” The experimenter drops the items together at the same time and asks: “Did you notice anything, what?”

“Did the same thing happen to both objects?” (description)

Experimenter: “Okay thank you. Why do you think things happened that way?”

“Do you think there might be another reason for that?” (explanation)

For *absorption*, children saw water rising from a petrie dish through strips of tissue and blotting paper of the same length/width, the water rising faster through the more open structure of the tissue. Similar to sinking protocol, children needed to predict outcomes
ahead of witnessing simultaneous demonstration of the two instances, then to describe e.g. the rates of absorption, and to explain the outcomes, as a measure of causal inference assessing identification of basic factors (e.g. softness/hardness of the materials), operative variables (e.g. relative softness/hardness/porousness of the materials) and mechanisms (e.g. relative size of the holes allowing water to rise up). A typical testing session involved the following scripts:

Experimenter: “I have got these two strips of paper here: a tissue and a blotting paper. Do you want to touch them?”

“I am going to dip them in the water. What do you think will happen?”

“Do you think that the same thing will happen to both papers?” (predictions)

Experimenter: “Okay. Please watch carefully now. I will dip the papers and you will watch.” The experimenter dips the end of the papers in water together at the same time and asks: “Did you notice anything, what?”

“Did the same thing happen to both papers?” (descriptions)

Experimenter: “Why do you think things happened that way?”

“Do you think there might be another reason for that?” (explanation)

For solution, children saw the same small quantities of table and rock salt dissolve in warm water. The small quantity of the salt was assured with two equally very small spoons. The greater surface area to volume of the table salt led to more rapid solution (see Figure 2, Chapter 2). Similar to sinking and absorption protocols, children needed to predict outcomes ahead of witnessing simultaneous demonstration of the two instances, then to describe e.g. the rates of absorption, and to explain the outcomes, as a measure of causal inference assessing identification of basic factors (e.g. softness/hardness of the materials), operative variables (e.g. relative softness/hardness/compactness of the materials) and mechanisms (e.g. relative size of the grains allowing water to penetrate in).

A typical testing session involved the following scripts:

Experimenter: “I have got these two kinds of salts here: table salt and rock salt. I am going to drop a tiny piece from each in the water. What do you think will happen?”

“Do you think that the same thing will happen to both kinds of salt?” (predictions)
Experimenter: “Please watch carefully now. I will drop the salts and you will watch.” The experimenter drops the samples of the salts together at the same time and asks: “Did you notice anything, what?”

“Did the same thing happen to both kinds of salt?” (descriptions)

Experimenter: “Why do you think things happened that way?”

“Do you think there might be another reason for that?” (explanation)

Each task had the same three-stage structure, in which children: (i) inspected the contrasting materials and were asked what they thought would happen when they were put into the water (prediction from prior knowledge); (ii) watched the focal events and were asked to describe what they had noticed; (iii) were asked to explain why they thought things had happened in the way that they had seen. At each stage, they were encouraged to give as full an answer as they could. Observation duration was longer for dissolving due to the salt types took some reasonable amount of time to dissolve. Observation duration was shorter for sinking, and moderate for absorption.

Data from these tasks were used to compute three types of measures: (1) components: individual total scores for accurate prediction from prior knowledge, description, and explanation across three tasks, (2) composite scores for each task (sinking, absorption, solution), and (3) a total causal score combining these.

(1) Components: children’s responses for prior knowledge, description and explanation computed across three tasks (Table 7 for the scoring system). Prior knowledge and description were scored for accuracy of anticipating / reporting differences in sinking/absorption/solution rate (1 point per object. Therefore the scores were between 0-6 for both prior knowledge and description). Explanation scoring began at the minimal level of the observed factor(s) (score of 1); via making explicit that these are variables linked to the observed differences in speed of the contrasting examples (score of 2); to a statement about the underlying mechanism which produced the effect (score of 3). Therefore the scores were between 0-9) across the three tasks.
(2) For the composite measure, children’s responses for each phenomenon scored independently. The minimum and maximum score for each phenomenon varied from 0 to 7.

(3) For the total causal score all these scores were combined (0-21; alpha = .751). The total number of mechanism level explanation responses were also noted across the tasks (0-3), as a separate measure of higher causal thinking. Table 7 provides a detailed version of the scoring system.

Table 7. Scoring system for causal tasks

<table>
<thead>
<tr>
<th>Component</th>
<th>Sinking</th>
<th>Absorption</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction from prior knowledge</td>
<td>Correct prediction for stone (i.e. sinks) =1</td>
<td>Correct prediction for tissue paper =1</td>
<td>Correct prediction for table salt =1</td>
</tr>
<tr>
<td>(0-2)</td>
<td>Correct prediction for difference between stone and berry (i.e. sink at different speeds) =1</td>
<td>Correct prediction for difference between tissue and blotting paper =1</td>
<td>Correct prediction for difference between table and rock salt =1</td>
</tr>
<tr>
<td>Description of observation</td>
<td>Correct description for stone =1</td>
<td>Correct description for tissue paper =1</td>
<td>Correct description for table salt =1</td>
</tr>
<tr>
<td>(0-2)</td>
<td>Correct description for berry =1</td>
<td>Correct description for blotting paper =1</td>
<td>Correct description for rock salt =1</td>
</tr>
<tr>
<td>Explanation/inference</td>
<td>No/irrelevant explanation =0</td>
<td>No/irrelevant explanation =0</td>
<td>No/irrelevant explanation =0</td>
</tr>
<tr>
<td>(0-3)</td>
<td>Weight/size without difference between objects =1</td>
<td>Thickness/softness/texture etc. without difference between types of paper =1</td>
<td>Grain/size etc. without difference between types of salt =1</td>
</tr>
<tr>
<td></td>
<td>Weight/size with difference =2</td>
<td>Thickness/softness/texture etc. with difference =2</td>
<td>Grain/size etc. with difference =2</td>
</tr>
<tr>
<td></td>
<td>Density and</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To confirm reliability, two researchers subsequently scored all responses independently from the audio-recordings. Agreement was 93%, and final scores were assigned following discussion and checking the audios in the small number of instances where there was a difference (example responses are shown in the Table 8).

### Table 8. Examples of explanation responses

<table>
<thead>
<tr>
<th>Phenomena</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinking</td>
<td>“They are heavy and they sank to the bottom”</td>
<td>‘The stone is heavier than the berry so they sank to the bottom differently”</td>
<td>“They are both heavier than the water and cannot hold air in it so they sank to the bottom. But the stone sank quicker than the berry because it’s got more stuff in it so the water can’t hold it up as it did to berry.”</td>
</tr>
<tr>
<td>Absorption</td>
<td>“If you dip the paper in the water they get wet because they’re soft”</td>
<td>“The tissue paper is thinner than the other paper so water rises faster in it”</td>
<td>“The tissue paper has holes in it that help water to rise up. Water holds on the walls of the holes and layers and that helps it to climb up. Other paper has some space in it, but not as much as the tissue paper.”</td>
</tr>
<tr>
<td>Solution</td>
<td>“They go into water because they’re small”</td>
<td>“The table salt is smaller than the rock salt so it disappears”</td>
<td>“The size of the two types of salt is different. And this is more rocky so water cannot</td>
</tr>
</tbody>
</table>
and spread quicker.” go into it easily. They both dissolve in the water, but rocky one takes more time than the table salt.”

3.4.3.3 Measures of spatial-temporal analysis

3.4.3.3.1 The flow of liquid task
This task was adapted from Piaget (1969/2006), and examined children’s ability to analyze the liquid flow from one container to another at successive time points, and reconstruct the sequence of change. It consisted of three stages. At the first, two flasks were presented one on the top of the other with a tap between (Figure 4). The upper flask (I) was filled with red-coloured water, while the lower (II) was empty.

![Flasks at the start of the flow of liquid task](image1)

Flasks at the start of the flow of liquid task

Transferring of the liquid movement onto proformas

![Transferring of the liquid movement onto proformas](image2)

Figure 4. Flow of liquid task

Children were given a proforma showing both flasks with a space between them and, they marked the respective levels in the flasks by drawing horizontal lines on the proforma. The liquid was then allowed to flow from I to II in four further steps, and the child marked the liquid level on a fresh proforma each time, being invited to correct any errors.
At stage two, the proformas were shuffled and the child put them in order, again being invited to correct any errors. At the third stage, each proforma was cut in two, separating drawings of I from II, shuffled, and the child attempted to put them in order again. Scores were based on the number of drawings in the correct position at this stage, and could therefore range from 0-10. A typical testing session involved the following scripts:

First stage

Experimenter: “We have got two bottles here. The upper bottle is filled with red liquid; the lower bottle is empty at the beginning. When I open the tap, there will be two motions: a drop of level from here (showing upper bottle) and a rise of level here (showing lower bottle) like this…”

(Demonstration).

Experimenter: “Here are the pictures of the bottles.” [Before any water flows] “Can you draw a line showing the liquid level at the upper bottle (here)” (Waiting child to draw the line) “And lower bottle (here)?” (Waiting child to draw)

Experimenter: “Now I will allow some liquid to flow down to the bottom bottle. You see the liquid level has changed at the bottles. Can you draw a line showing the liquid level at the upper bottle (here) and lower bottle (here)?” (Waiting child to draw the lines)

Experimenter: “Now, I will open the tap again. The liquid level will change in both bottles. Can you draw a line showing the liquid level at the upper (here) bottle and lower (here) bottle?”

Experimenter repeated the conversation one more time

Experimenter: “Now, last time, I will open the tap again.” [Let water run completely to the bottom] Observe the liquid level at both bottles. “Can you draw a line showing the liquid level at the upper (here) bottle and lower (here) bottle?”

Note that if children made obvious errors in their drawings, the experimenter invited them to correct these on a fresh paper sheet. Recorded the number of errors, and time taken in seconds for this stage. However, these were not included in the scoring system. Extra care was taken for all children to have approximately the correct drawings. When it
was necessary, children received guidance at this stage, as their drawing ability was not the focus.

Second stage

Experimenter: “Look, we have five different pictures showing the drop and rise levels of the liquid.” Child’s drawings are shown to her/him with the mixed order on the table. “Can you put them in order until the upper bottle is emptied? Which one is your first drawing, second, third…?” At this stage, child’s performance is not scored but the numbers of errors are recorded. If the child made errors, this explained to child by the experimenter and she/he was invited to correct those.

Experimenter: “Great! Now, we will shuffle the pictures, and put them in order like we did before.” If the child made errors, she/he was invited to correct these. The number of errors were noted, and time taken in seconds for this stage.

Third stage

Experimenter: “Great! Now. I will separate upper bottle from lower bottle by cutting.” Experimenter cut the drawing from the middle, shuffled. “Can you put them in order like we did before?” At this stage, child’s performance is scored based on the number of drawings in the correct position. Children’s strategies to reconstruct the sequence were observed and noted.

3.4.3.3.2 The speed task

Children saw computer animations of three bunnies (red, yellow, black; colours were chosen on the basis of their resolution characteristics on the computerized screen) racing towards a carrot from different start positions at different speeds (Figure 5), with the animation stopping before they reached it. Children judged which bunny would arrive at the carrot first. They were requested to estimate the winner by watching the first half of the run. The task began with two practice items, followed by 17 trials gradually increasing in difficulty: the stop time reduced, from 6 to 4 seconds, as did the difference between the three bunnies in start point and relative speed, making differences in arrival
time harder to distinguish, and the period available within which to track the differences shorter. The number of correct responses was recorded (0-17). A typical testing session involved the following scripts:

Practice trials

Experimenter: “Now, you will see three bunnies with different colours: red, yellow, black. They will be racing for the carrot, because they are all very hungry. You will need to guess which bunny will reach to the carrot depending on its speed and distance. … Look at these examples…”

The child sees two practice items.

Experimenter: “There will be 17 different races, and each time please tell me the colour of the winner bunny, so I can tell you how well you observed the entire race. Shall we get started?”

Based on the child’s response actual trials started. Each time child was asked to tell the colour of the winner bunny. At the end, children were not given feedback about the winner bunnies, but all children were congratulated and thanked without any exception (e.g. “Great job, well done. Thank you”).

Figure 5. Example configuration of the bunnies at the start of a speed task trial
3.4.3.4 Measures of spatial analysis

3.4.3.4.1 The monkey mental rotation
This task was adapted by Broadbent, Farran and Tolmie (2014) from Shepard and Metzler’s (1971) study. Children saw laminated sheets with two mirror image cartoon monkeys above a horizontal line and one below at varying degrees of rotation from upright (Figure 6a), and chose which of the two top monkeys matched the one underneath. Three practice trials were followed by 16 experimental trials (2 x 0° trials; 4 x 45° trials, 4 x 90° trials, 4 x 135° trials, with equal numbers of clockwise and anticlockwise rotations of each target monkey; and 2 x 180° trials), presented in pseudorandom order. The number of correct responses was recorded (0-16). A typical testing session involved the following scripts:

Practice trials
Experimenter: “Now you will see two cartoon monkeys on the top of this horizontal line. One monkey below the line is rotated in several ways.” The experimenter shows the first practice trial. “I will ask you to choose which of the two monkeys on the top (here) matched the one underneath (here). Please be careful that both monkeys at the top had one red hand pointing upwards above their tails (here and here). Blue hands (here and here) pointing downwards. The incorrect monkey is always a mirror image of the correct one. You will need to rotate the monkey in your head and tell me the right answer.”

Experimenter: “Lets take a look at that one.” Introducing another practice laminated sheet. “You see, the monkey is rotated differently now. Can you match the monkey underneath (here) with the same one at the top? Is it the same as this or this one?”

Children’s responses were discussed with them. If they did not understand the target, the explanation was repeated. If they made errors, they were invited to correct it with the guidance of the experimenter.

Experimenter: “Great! Let’s take a look at the last practice item.” The same explanations
and question were repeated. “Great! If you are ready, shall we start to look at the rest?”

Actual trials started after the child’s response. Children did not receive any guidance or feedback about their performance. At the end, they were all congratulated and thanked.

3.4.3.4.2 Paper-folding
This task was adapted by Harris et al. (2013) from Eliot and Smith’s (1983) study. Children saw laminated coloured sheets of paper with markings indicating a fold, and selected from four options the one that represented the outcome (Figure 6b). There were two practice items, and 14 trials, which increased in difficulty by employing more complex shapes and folds, and response options that were harder to distinguish from each other. The score was the number of correct responses (0-14). A typical test session involved the following script:

Practice items
Experimenter: “Please look at this piece of paper at the top (here). See? It is green on the front (here). If we turn it over the background is purple on the other side (here). These arrows (here) show us how to fold the paper. We fold the edge where the arrow starts (here) to where the arrow points. These dotted lines show us where the fold should be.
Let’s see how to guess the answer: if we fold the paper here how does it look like? Does it look like this, … or this one at the bottom?”

Children’s answers were discussed with themselves (e.g. why the answer is correct/wrong when the corners are folded, is it on the right or wrong side, thinking about colour to guide the estimation etc.) during the first two practice trials.

Similarly, actual trials started after the child’s response. Children did not receive any guidance or feedback about their performance. At the end, they were all congratulated and thanked.

3.4.3.5 Statistical thinking tasks
The Randomness task was used to explore children’s understanding of uncertainty and chance. Participants were shown two identical decks of 30 cards, five of which had smiley face stickers, with the remainder blank. The cards with the stickers were placed at the top of each deck; face up, so that they were visible. One of the decks was then shuffled so that the cards with smiley faces were now mixed with the blank cards. The two decks were then put face down, and participants were asked; “If you want to make sure to pick a smiley face, which deck would you pick from?” Children’s choices were marked as 0 or 1 depending on whether they chose the shuffled or unshuffled deck. Then, children were asked; “Could you explain why did you choose from this deck?” Regardless of they made the correct choice all children were posed the second question. Their explanations were marked as 0 or 1 according to whether they were able to identify the predictability of the position of the cards with the smiley faces as key to making a choice. Scores could range from 0 to 2.

The Marbles task used to explore children’s understanding of proportional calculations of probability. This task was adopted from Piaget and Inhelder (1975) whom used the marbles to test children’s proportional calculations and probability judgments. Children were shown four trays with different numbers of bicolored marbles over four trials (see Figure 7a). After being told that blue marbles were the winners, children were asked
“Imagine that I will ask you to close your eyes and pick one marble form each tray one at a time. While you pick, I will be tilting the tray to make sure that you cannot see the colour of the marble. You will tell me how good each tray is for winning?” The experimenter started from the first tray, put it closer to the child by separating from the others and asked “Okay, let’s take a look at the first tray. What do you think your chance is to pick a winner from that tray?” Children were also asked “How good is this tray for winning?” Each tray was introduced after child answered. After each question, they are requested to answer verbally as either fractions/ratios (as some older children did spontaneously), or by ticking on a line on the scoring sheet starting from ‘never get one’ to ending with ‘always get one’. Fully correct answers on both parts of the question were scored as 2 points for each tray, and partially correct ticking scored as 1. Participants who gave consistent correct answers for the second and the fourth tray received an extra 2 points for confirming verbally that the proportions were identical. This yielded an overall score ranging from 0 to 10.

(a) Marbles task with the trial order

![Marbles](image1)

(b) Cards (the first deck)

![Cards](image2)

Figure 7. (a) The four trays for the four trials of the Marbles task, in trial order; (b) the deck shown in the first trial of the cards task - only one smiley and one frown were dealt face up, the other two cards were shown face down
The *Cards task* assessed understanding of frequentist probability developed by the researchers. Children saw over four trials four decks comprised of different numbers of cards with smiley versus sad face stickers (see Figure 7b). Four trials respectively contained (1) two smiley, two sad; (2) two smiley, four sad; (3) four smiley, two sad; (4) four smiley, four sad cards, thus utilising the same proportions as in the marbles task, to ensure that any differences in difficulty between marbles and card tasks did not just reflect differences between samples presented. On each trial, they saw half of the cards dealt out face up, selected to represent the overall proportions, with the others remaining face down. Children were told “Okay, now you will see four decks of cards, each has different amount of smiley and sad faces like this. I will show you the half of the deck and ask you to guess how good each deck is for picking a smiley. And then like marbles you will need to estimate your chances of doing so. Ready? Okay. This is the first deck; you see one smiley, one sad face. What do you think your chance would be to pick a smiley from this deck?” The question was repeated for next three decks. In the second deck, children saw one smiley, two sad faces; in the third, they say two smiley one sad face; and at the end, they saw two smiley two sad faces. Scores were as for the marbles task, and could range from 0 to 10.

The *Covariation task* involved three trials on a laptop, each displaying, in pseudorandom order, a series of 8 pictures, half in a square, half in a circle frame. Children had to detect whether there was a relation between frame shape and content of the picture.

The task was started with an introduction, displaying a triangle accompanying with a flower (Figure 8). Children were told “Now, you are going to see some shapes appear one by one on the screen filled with different pictures like this (the child is presented with the figure 8): a flower goes with a triangle. I will ask you each time to look at the screen carefully and tell me what shape goes with what picture.

Figure 8. Practice display of the covariation task
The first display showed perfect covariation, four pictures of an ice cream in the square and four of a star in the circle (see Figure 9). Each figure appeared on the screen one by one, Children were asked: “Which shape goes with the ice cream?”, “Do they always go together?”, “Which shape goes with the star?”, and “Do they always go together?”

The second display showed imperfect covariation (75% contingency): three pictures of a basketball and one of sunglasses in the squares, and three of a phone and one of a line in the circle. Participants were asked: “Which picture goes with the circle?”, “Do they always go together?”, “Which picture goes with the square?”, and “Do they always go together?” The third display had no pattern (zero contingency), the circles and squares all contained different pictures, and participants were again asked the same questions. Shapes were kept constant to provide a common anchor across displays, but pictures were varied to avoid carry-over. All trials consisted of 8 figures each. Co-occurrence could be expressed as fractions/ratios, or by ticking on a line, as for marbles, from ‘can’t tell at all’ to ‘definitely’. Each correct answer was marked as 1 point. Children were expected to identify of the dominant correlate for displays 1 and 2, and they were supposed to say ‘none’/‘any’ for display 3 based on the appropriate estimation of the strength of association. Scores could range from 0 to 12.
3.5 Results

Analyses utilised data from all 107 participants who completed testing. Two-sided tests were used where relevant, with p<.05. The observed power for regression analyses was 0.95. One-way-ANOVAs were used to observe significant differences between the means. Zero-order and partial correlations illustrated the associations. Frequency distributions, skewness, and trends were considered. Further, hierarchical regression and maximum likelihood approaches were employed to examine the unique variances. The mediation and moderation analyses were employed to distinguish between the effects and interactions where necessary.

3.5.1 Developmental trajectories

3.5.1.1 Causal tasks

*Composite scores.* Figure 10 shows the response profiles for each age group on the sinking, absorption and solution tasks, based on composite scores across. Performance was best on sinking, followed by absorption, with solution some way behind.

A two-way mixed ANOVA (task within-subjects, age between-subjects) found a significant main effect of task, F(2,208)=47.202, p<.001, partial eta-squared=.312, with no significant difference between scores on sinking and absorption using a Bonferroni comparison, but with both significantly higher than scores for solution. There was also a main effect of age group, F(2,104)=24.250, p<.001, partial eta-squared=.318, with scores for Y1 significantly lower than Y3, but no difference between Y3 and Y5; and a modest task x age interaction, F(4, 208)=5.056, p=.001, partial eta-squared=.089, reflecting greater growth on solution between Y1 and Y3 than for sinking and absorption.
Figure 10. Profile of scores on (a) sinking, (b) absorption, (c) solution (max=7)
Components and causal total scores. The profiles of each age group for prior knowledge, description and causal explanation across the three causal tasks are shown in Figure 11. Children performed at a high level on description, slightly less well on prior knowledge, and at a notably lower level on causal explanation. For description, 91.6% of children obtained the maximum score in the sinking task, 85% in absorption, and 57.9% in solution. For prior knowledge, the corresponding values were 72.9% for sinking, 62.6% for absorption, and 37.4% for solution.

In contrast, for explanation, only 9.3% got the highest score in sinking, 14% for absorption, and 4.7% for solution. The majority of explanation responses on all three tasks focused solely on identification of causal factors or variables (scores of 1 or 2). Although mechanism responses became more common in the two older groups (for sinking, there were 0 in Y1, 4 in Y3 and 5 in Y5; for absorption, 1, 6 and 8 respectively, and for solution, 0, 0 and 5), children apparently found it difficult to make the shift to this level of thinking. However, if they made any reference to mechanism at all, they tended to do so on more than one task, at over 2.5 times the chance rate, so the shift appeared to be domain-general when it occurred.

Direct statistical comparisons between the components were not made, given differences in the scales and dimensions measured. However, one-way ANOVAs showed age-related progression on each: for prior knowledge, F(2,104)=12.376, p<.001, partial eta squared=.192; for description, F(2,104)=18.336, p<.001, partial eta squared=.261; for explanation, F(2,104)=17.383, p<.001, partial eta squared=.251. For each component, there were significant differences between Y1 and Y3, but not between Y3 and Y5. The components were positively correlated with each other, controlling for age (for prior knowledge and description, r=.392; for prior knowledge and explanation, r=.414; for description and explanation, r=.618, all p<.001).
Figure 11. Profile of scores on (a) prior knowledge (max=6), (b) description (max=6), causal explanation (max=9)
Children’s descriptions were close to the ceiling, but they performed at lower level on prior knowledge and in particular on explanation. For the three main components, there were significant differences between Y1 and Y3, but not between Y3 and Y5; for mechanisms, there was a significant difference between Y1 and Y5, indicating later growth for responses at this level (see Table 9).

There was a clear distinction between the profiles for sinking, absorption, and solution. Children found solution more difficult in this sample. However, age was not the only determinant of performance. There were also some signs of a disjunction between better and poorer performers, especially in Y1 for sinking, and in Y5 for solution.

Table 9. Mean score (sd) by age group on total causal score (max=21), prior knowledge, description (max=6), explanation (max=9), and mechanism (max=3)

<table>
<thead>
<tr>
<th></th>
<th>Y1</th>
<th>Y3</th>
<th>Y5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal total</td>
<td>10.63 (4.44)</td>
<td>14.42 (2.96)</td>
<td>15.97 (2.44)</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>3.26 (1.52)</td>
<td>4.15 (1.50)</td>
<td>4.85 (1.09)</td>
</tr>
<tr>
<td>Description</td>
<td>4.03 (1.60)</td>
<td>5.21 (0.99)</td>
<td>5.59 (0.68)</td>
</tr>
<tr>
<td>Explanation</td>
<td>3.34 (1.89)</td>
<td>5.06 (1.54)</td>
<td>5.54 (1.54)</td>
</tr>
<tr>
<td>Mechanism</td>
<td>0.03 (0.17)</td>
<td>0.30 (0.58)</td>
<td>0.44 (0.72)</td>
</tr>
</tbody>
</table>

The majority of inference responses on all three causal tasks focused solely on relevant factors or variables (scores of 1 or 2). Mechanism responses were more apparent among older children: 2.9% of children in Y1 gave one or more mechanism response, 24.2% in Y3, and 30.8% in Y5. The details for the response profiles of the causal tasks can be found in Chapter 3.
3.5.1.2 Spatial-temporal measures

Response profiles for each age group on the flow of liquid and speed tasks (see Figure 12), and one-way ANOVAs showed age-related increases on both, but growth occurring later for speed: for flow of liquid, F(2,104)= 11.180, p<.001, partial eta squared=.177, with significant differences between Y1 and Y3, but not between Y3 and Y5; and for speed, F(2,104)=16.492, p<.001, partial eta squared=.241, with no difference between Y1 and Y3, but both significantly different from Y5.

For flow of liquid, there was steady growth with age in the percentage of children obtaining perfect scores of 10, with a fading tail showing children who made one or more errors. In Y1, 17 children (48.6%) exhibited a perfect performance; scores otherwise ranged from 0 to 8 and more or less normally distributed across this range. There was a comparable pattern in Y3 and Y5, where 22 (66.7%) and 35 children (89.7%) respectively obtained perfect scores, and other scores fell between 4 and 8.

Looking at children’s performance across all stages of the task, 20 (18.7%) failed to transfer liquid levels onto paper appropriately at stage 1, while 14 (13.1%) failed to put their drawings in sequential order at stage 2. These children also tended to make mistakes at stage 3: errors correlated at .24, p<.05, for stage 1 and 2; .35, p<.001, for stage 1 and 3; and .20, p<.05, for stage 2 and 3. Failure at stage 1 in particular was a precursor of failure at stage 3. Taken together, the data suggest a distinction between a growing number of children with age who had a perfect strategy for the task; and a declining number who failed to grasp it, and who made often multiple errors.

Performance on the speed task suggested a dichotomy between better and poorer ability. For Y1, scores ranged from 6 to 16, showed normal distribution, with frequency peaking at 11. The mean for Y3 was virtually identical, but the distribution was different, with suggestions of bimodality. In total, 10 children (30.4%) fell into the upper range; 2 children obtained a perfect score of 17. There was a similar pattern for Y5. Scores ranged from 8 to 17, but with separate peaks at 10 to 11, and at 14. For Y5, scores ranged from 9 to 17, but separate peaks were observed at 11 and at 17, and a tendency towards ceiling for the upper group.
Figure 12. Profile of scores on (a) flow of liquid (max=10), and (b) speed (max=17)
Overall, in spite of different item properties, scoring systems and developmental trends, flow of liquid and speed both suggested the existence of a disjunction between children with a good grasp of task requirements, who increased in number with age, and children who made errors.

3.5.1.3 Verbal, nonverbal and spatial ability
There was significant positive skew on block design, due to the oldest age group having a longer tail. Vocabulary was normally distributed. One-way ANOVAs found significant increases with age on both, however: for vocabulary, Welch robust statistic=54.093 (df = 2, 67.790); for block design, 45.070 (2, 63.948), p<.001 for both, with significant differences between all three age groups on both measures: for vocabulary, Y1 mean=22.89, sd=5.290, Y3 mean=30.76, sd=5.863, Y5 mean=35.62, sd=5.204; for block design, Y1 mean=11.91, sd=6.085, Y3 mean=19.15, sd=9.517, Y5 mean=34.19, sd=13.250. Variance was not notably decreased for either measure: for vocabulary, overall mean=29.95, sd=7.586; for block design, overall mean=22.23, sd=13.860.

Table 10 shows the age profile of performance on the spatial, verbal, and nonverbal measures. There was significant negative skew on rotation and paper folding, and positive on block design, due to the youngest group having a longer tail on the negatively skewed variables, and the oldest on the positive. Vocabulary was normally distributed.

Table 10. Mean score (sd) by age group on rotation (max=16), paper folding (max=14), vocabulary, and block design

<table>
<thead>
<tr>
<th></th>
<th>Y1</th>
<th>Y3</th>
<th>Y5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation</td>
<td>12.89 (3.89)</td>
<td>14.15 (2.42)</td>
<td>15.64 (0.58)</td>
</tr>
<tr>
<td>Paper folding</td>
<td>5.03 (3.59)</td>
<td>7.97 (3.04)</td>
<td>11.28 (1.89)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>22.89 (5.29)</td>
<td>30.76 (5.86)</td>
<td>35.62 (5.20)</td>
</tr>
<tr>
<td>Block design</td>
<td>11.91 (6.08)</td>
<td>19.15 (9.52)</td>
<td>34.10 (13.25)</td>
</tr>
</tbody>
</table>
One-way ANOVAs found significant increases with age on each: for rotation, F(2,104)=13.901, p<.001; for paper folding, F(2,104)=47.983, p<.001; Welch robust statistics for rotation=13.901 (df=2, 46.893); for paper folding=47.983 (df=2, 60.628); for vocabulary=54.093 (df=2, 67.790); for block design, 45.070 (df=2, 63.948); and p<.001 for all. For rotation, there were significant differences between Y1 and Y5; for the other measures, all three age groups differed significantly. Progress on the spatial tasks was therefore slightly later than for flow of liquid, where there were no differences between Y3 and Y5; but earlier than for speed, where there were no differences between Y1 and Y3. However, performance on rotation began to approach ceiling by Y3, in common with flow of liquid.

3.5.1.4 Statistical tasks
There was significant negative skew on randomness, marbles and cards. Covariation was normally distributed. Figure 13 compares the developmental trajectory for each measure using scores standardized to a scale between 0 and 1 for comparability.

![Figure 13. Developmental trajectories](image)

Overall, there was a clear upward trend for all tasks, confirming the age effect, but with variation in relative difficulty (see also Table 11). Blocks task was in particular difficult for most children, while flow of liquid (FOL) was easier (hence the difference in direction of skew), with causal total lying in between. Comparing the trends, the steepest
gradients were seen for blocks and covariation, followed by randomness, cards and marbles.

In terms of differences between the age groups, marbles showed a different pattern: the steepest increase was between Y1 and Y3, with a slow down subsequently. For the rest of the statistical tasks, the steepest gradient was between Y3 and Y5, except randomness where growth was linear.

Table 11. Mean score (sd) on randomness (max=2), marbles, cards (max=10), covariation (max=12)

<table>
<thead>
<tr>
<th></th>
<th>Y1</th>
<th>Y3</th>
<th>Y5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomness</td>
<td>1.09 (0.89)</td>
<td>1.42 (0.79)</td>
<td>1.90 (0.38)</td>
<td>1.49 (0.78)</td>
</tr>
<tr>
<td>Marbles</td>
<td>4.51 (3.08)</td>
<td>7.45 (2.95)</td>
<td>9.26 (1.82)</td>
<td>7.15 (3.29)</td>
</tr>
<tr>
<td>Cards</td>
<td>4.69 (3.56)</td>
<td>7.15 (2.76)</td>
<td>9.10 (1.59)</td>
<td>7.06 (3.27)</td>
</tr>
<tr>
<td>Covariation</td>
<td>4.83 (2.18)</td>
<td>6.61 (3.29)</td>
<td>9.82 (2.81)</td>
<td>7.20 (3.48)</td>
</tr>
</tbody>
</table>

One-way ANOVAs by school year found highly significant increases with age on all variables, using the Welch robust statistic, p<.001 in each case.

3.5.2 What predicts children’s reasoning of causal processes?

3.5.2.1 Correlations between variables
Zero-order Pearson correlations showed total causal score was positively associated with all the potential predictors, which all positively correlated with each other (Table 12). The relationship of the predictor variables to causal performance was linear, apart from block design, which was logarithmic, R-square for linear fit=.263; R-square for logarithmic fit=.368.

When age in months, vocabulary and logarithmic block design were controlled for, only
flow of liquid, marbles, cards and covariation remained significantly associated with causal score, as well as components (except cards). Speed was unrelated to any other variable, including flow of liquid. Parental occupation and education correlated with each other, $r=.609$, but otherwise only with nonverbal ability, $.261$, $p=.008$ and $.382$, $p<.001$ respectively, and are not considered further.

Table 12. Zero-order and partial correlations (zero-order above diagonal, N=107; partial correlations below diagonal, controlling for age in months, verbal and nonverbal ability,

<table>
<thead>
<tr>
<th></th>
<th>Causa</th>
<th>Inference</th>
<th>Vocabulary</th>
<th>LNblock</th>
<th>FOL</th>
<th>Rand</th>
<th>Marble</th>
<th>Cards</th>
<th>Covariation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal total</td>
<td>1</td>
<td>.90***</td>
<td>.54***</td>
<td>.61***</td>
<td>.52***</td>
<td>.39***</td>
<td>.55***</td>
<td>.52***</td>
<td>.56***</td>
</tr>
<tr>
<td>Prior</td>
<td>.69***</td>
<td>.53***</td>
<td>.47***</td>
<td>.56***</td>
<td>.46***</td>
<td>.27**</td>
<td>.42***</td>
<td>.42***</td>
<td>.47***</td>
</tr>
<tr>
<td>Description</td>
<td>.79***</td>
<td>.70***</td>
<td>.44***</td>
<td>.48***</td>
<td>.45***</td>
<td>.40***</td>
<td>.42***</td>
<td>.49***</td>
<td>.43***</td>
</tr>
<tr>
<td>Inference</td>
<td>.85***</td>
<td>1</td>
<td>.47***</td>
<td>.52***</td>
<td>.42***</td>
<td>.34***</td>
<td>.53***</td>
<td>.43***</td>
<td>.51***</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>.68***</td>
<td>.44***</td>
<td>.44***</td>
<td>.52***</td>
<td>.53***</td>
<td>.64***</td>
</tr>
<tr>
<td>LNblock</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>.43***</td>
<td>.41***</td>
<td>.56***</td>
<td>.54***</td>
<td>.62***</td>
</tr>
<tr>
<td>FOL</td>
<td>.33**</td>
<td>.22*</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>.35***</td>
<td>.55***</td>
<td>.56***</td>
<td>.47***</td>
</tr>
<tr>
<td>Speed</td>
<td>.07</td>
<td>.04</td>
<td>-</td>
<td>-</td>
<td>.25*</td>
<td>.26*</td>
<td>.42***</td>
<td>.40***</td>
<td>.59***</td>
</tr>
<tr>
<td>Paper fold</td>
<td>.14</td>
<td>.06</td>
<td>-</td>
<td>-</td>
<td>.43***</td>
<td>.45***</td>
<td>.50***</td>
<td>.52***</td>
<td>.59***</td>
</tr>
<tr>
<td>Rotation</td>
<td>.17</td>
<td>.16</td>
<td>-</td>
<td>-</td>
<td>.20*</td>
<td>.31*</td>
<td>.48***</td>
<td>.43***</td>
<td>.39***</td>
</tr>
<tr>
<td>Randomness</td>
<td>.13</td>
<td>.10</td>
<td>-</td>
<td>-</td>
<td>.35**</td>
<td>1</td>
<td>.49***</td>
<td>.54***</td>
<td>.43***</td>
</tr>
<tr>
<td>Marbles</td>
<td>.25*</td>
<td>.28**</td>
<td>-</td>
<td>-</td>
<td>.38***</td>
<td>.28**</td>
<td>1</td>
<td>.76***</td>
<td>.63***</td>
</tr>
<tr>
<td>Cards</td>
<td>.21*</td>
<td>.13</td>
<td>-</td>
<td>-</td>
<td>.38***</td>
<td>.36**</td>
<td>.61***</td>
<td>1</td>
<td>.60***</td>
</tr>
<tr>
<td>Covariation</td>
<td>.20*</td>
<td>.21*</td>
<td>-</td>
<td>-</td>
<td>.20*</td>
<td>.15</td>
<td>.37***</td>
<td>.32**</td>
<td>1</td>
</tr>
</tbody>
</table>

Zero-order correlations above diagonal, N=107; partial correlations below diagonal, N=106 due to missing date of birth data for one participant; *$p<.05$, **$p<.01$, ***$p<.001$
3.5.2.2 Hierarchical regression models

Hierarchical regression was used to examine the unique variance accounted for by the predictors. These tell us whether the change in R2 is significant.

*Initial model:* taking total causal score as the dependent variable, age in months and WASI vocabulary were entered in the first stage of the analysis. The other control variable, log block design, was entered at the second stage, to assess its effects separately from those of vocabulary. Speed was included on its own at the third stage to check with its predictive value before flow of liquid was entered at the fourth stage. Three models were trialed to examine the impact of spatial measures. First, spatial measures were entered before the spatial-temporal measures at the third and fourth stages. Then they were entered after the spatial-temporal measures. The spatial measures did never survive in these models; they remained non-significant with the lowest beta values. Another model trialed the impact of the spatial measures further, in which both spatial measures were entered immediately after the vocabulary. This inclusion showed that they shared their variances with the WASI vocabulary from the beginning, rather than with nonverbal or spatial-temporal measures. Statistical measures were added at the final stage one by one.

Although a number of models tested different combinations, the outcomes consistently showed that randomness, speed, and paper folding were nonsignificant, with the lowest beta values. In these initial models, only flow of liquid (spatial-temporal measure) remained significant with a beta value of .217, $p=.023$.

*Final model:* following these trials, the three nonsignificant measures were excluded from the final model. Note that the partial correlations also showed that randomness, speed, and paper folding were least related to causal measure. The final model included 8 variables, allowing a regression for the sample of 107 (see Field, 2013). First three control variables (age and IQ measures) were entered before other variables, followed by the spatial-temporal and spatial measures. The statistical measures were entered separately at each stage, but the beta values did not change. At the end, they were included in the final model all together, as shown in Table 13.
This analysis produced significant models and R-square change at the first three stages. Age and vocabulary were significant predictors at the first stage, but they became non significant when log block design was included. Flow of liquid was a significant predictor when entered at the third stage, explaining 5.4% of the variance with a probability equal to .02.

Table 13. Hierarchical regression with total causal score as dependent variable (significant predictors in bold)

<table>
<thead>
<tr>
<th>Model</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictor</td>
<td>β</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age in months</td>
<td>.332**</td>
<td>.207</td>
<td>.159</td>
<td>.139</td>
<td>.106</td>
</tr>
<tr>
<td>WASI vocabulary</td>
<td>.310**</td>
<td>.135</td>
<td>.089</td>
<td>.076</td>
<td>.039</td>
</tr>
<tr>
<td>Block design (log)</td>
<td>.383**</td>
<td>.329*</td>
<td>.278*</td>
<td>.242*</td>
<td></td>
</tr>
<tr>
<td>FOL</td>
<td>.266*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotation</td>
<td>.156</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cards</td>
<td>.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marbles</td>
<td>.060</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covariation</td>
<td>.116</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AdjRsquare = .502; $\Delta R^2 = .347^{***}$ for M1; .071*** for M2; .054** for M3; .019 for M4; .012 for M5; *p <.05. **p<.01. ***p<.001.

The model showed flow of liquid and log block design as significant predictors only. Similarly, the logarithmic trend of block in predicting causal scores highlighted that the steeper gradient of relationship was at lower levels, indicating that nonverbal ability made most difference at lower level of causal thinking. The linear effect of flow of liquid performance indicates, in contrast, that this measure of spatial-temporal analysis distinguished lower from higher levels of causal performance. This ability appeared to be largely independent not just of nonverbal and spatial ability but of verbal ability too.
Further moderation and mediation analyses, using the Hayes (2013) method, confirmed the lack of interaction between flow of liquid and vocabulary in the prediction of causal scores; and that the relationship between flow of liquid and causal performance was primarily a direct one, not mediated by vocabulary.

These models were also run for the three causal components separately, and produced similar outcomes, block design and flow of liquid consistent predictors of all components, except for explanation, in which none of the variable survived in the model (see Table 14). One of the spatial measures, rotation was a significant predictor of description along with flow of liquid.

Table 14. Hierarchical regression with components as dependent variables (significant predictors in bold)

<table>
<thead>
<tr>
<th>Model</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictor</td>
<td>β</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age in months</td>
<td>.197</td>
<td>.059</td>
<td>.013</td>
<td>.010</td>
<td>-.005</td>
</tr>
<tr>
<td>WASI vocabulary</td>
<td>.330**</td>
<td>.136</td>
<td>.092</td>
<td>.091</td>
<td>.057</td>
</tr>
<tr>
<td>Block design (log)</td>
<td>.425***</td>
<td>.366*</td>
<td>.366*</td>
<td>.340*</td>
<td></td>
</tr>
<tr>
<td>FOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marbles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covariation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AdjRsquare = .380; ΔR² = .238*** for M1; .087*** for M2; .048 for M3; .000 for M4; .007 for M5. *p &lt; .05. **p &lt; .01. ***p &lt; .001.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description

<p>| Age in months | .379* | .300* | .255* | .226 | .220 |</p>
<table>
<thead>
<tr>
<th></th>
<th>WASI vocabulary</th>
<th>Block design (log)</th>
<th>FOL</th>
<th>Rotation</th>
<th>Cards</th>
<th>Marbles</th>
<th>Covariation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.178</td>
<td>.243*</td>
<td>.246*</td>
<td>.228*</td>
<td></td>
<td></td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>.067</td>
<td>.193</td>
<td>.251*</td>
<td>.221*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.025</td>
<td>.117</td>
<td>.223*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.006</td>
<td>.108</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AdjRsquare = .402; \( \Delta R^2 = .269^{***} \) for M1; .028 for M2; .046* for M3; .040* for M4; .019 for M5. *\( p < .05 \); **\( p < .01 \); ***\( p < .001 \). + The relationship of paper-folding to description was logarithmic.

**Explanation**

<table>
<thead>
<tr>
<th></th>
<th>WASI vocabulary</th>
<th>Block design (log)</th>
<th>FOL</th>
<th>Rotation</th>
<th>Cards</th>
<th>Marbles</th>
<th>Covariation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in months</td>
<td>.285*</td>
<td>.185*</td>
<td>.150</td>
<td>.130</td>
<td>.077</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.272*</td>
<td>.132</td>
<td>.098</td>
<td>.085</td>
<td>.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.307*</td>
<td>.268*</td>
<td>.216</td>
<td>.168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.194*</td>
<td>.197*</td>
<td>.135</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.156</td>
<td>.107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AdjRsquare = .391; \( \Delta R^2 = .262^{***} \) for M1; .045* for M2; .028*** for M3; .019 for M4; .036 for M5. *\( p < .05 \); **\( p < .01 \); ***\( p < .001 \).

Relationships between the three dependent measures and the predictors were all linear, except for block design, which was again consistently logarithmic. The predictors, with that adjustment for paper folding, were entered in the same order as for the overall regression analysis. These analyses revealed both commonalities and differences in the predictors of the causal component scores.
For prior knowledge, the final model, adjR-square=.380, had only log block design, beta=.340, p=.007, 8.7% change in explained variance, and flow of liquid, beta=.234, p=.025, 4.8% change in explained variance as significant predictors. In contrast, for description, the final model, adjusted adjR-square=.402, had rotation, beta=.221, p=.020, and 4% change in explained variance, and flow of liquid, flow of liquid, beta=.223, p=.030, with 4.6% change in explained variance, as significant predictors. For inference, although the first three models were significant, there was no significant predictor at the final stage. However, flow of liquid remained to be significant until statistical measures were included. This indicated they had overlaps in their shared variances.

All these models indicated that in this high achieving sample neither age, nor verbal, spatial, or statistical abilities were stable predictors of causal measures, instead nonverbal and spatial-temporal measures were more reliable.

3.5.2.3 Nature of shared variances between predictors

Factor analysis with varimax rotation (KMO=.834) was used to explore in more depth the nature of the relationship between flow of liquid, log block design, marbles, cards and covariation, given their shared influence on the causal measures. This identified a four-factor solution that explained 95% of the shared variance between the five measures, which confirmed separable components relating to marbles/cards, covariation, flow of liquid and log block design, as shown in Table 15.

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOL</td>
<td>.292</td>
<td>.927</td>
<td>.163</td>
<td>.168</td>
</tr>
<tr>
<td>Log blocks</td>
<td>.281</td>
<td>.172</td>
<td>.907</td>
<td>.264</td>
</tr>
<tr>
<td>Marbles</td>
<td>.821</td>
<td>.233</td>
<td>.234</td>
<td>.298</td>
</tr>
<tr>
<td>Cards</td>
<td>.856</td>
<td>.257</td>
<td>.223</td>
<td>.208</td>
</tr>
<tr>
<td>Covariation</td>
<td>.335</td>
<td>.189</td>
<td>.287</td>
<td>.876</td>
</tr>
</tbody>
</table>
3.5.2.4 Path analysis
A maximum likelihood approach was employed to test the fit of the regression model for causal score, treating age as a background influence on the predictors. The strongest predictors from the correlations and regressions were included. The best fit was provided by an extended mediation model: for total causal score, $\chi^2 = 3.891$, $p = .273$; for inference, $\chi^2 = 3.887$, $p = .274$; df = 3 for both. Figure 14 illustrates the model plus path coefficients obtained for total causal score.

![Figure 14](image)

Figure 14. Extended mediation model for the effects of log block design, covariation, marbles and flow of liquid on causal total score (subsidiary effects in grey)

There seems to be a stable pattern of effects in which nonverbal ability, awareness of covariation and probability (as indexed by marbles) support spatial-temporal analysis, but each influencing causal total to different degrees. Nonverbal and spatial-temporal ability have the largest direct effects, with the effects of probability and covariation smaller by comparison (for inference, the direct effect of probability, .22, was stronger than nonverbal and spatial-temporal ability, .19, .10, respectively). Age and vocabulary have
little or no direct impact on causal thinking in these models, and act as background variables, influencing the main predictors to different degrees.

Three models were tested in the path analyses: (a) an independent effects model, in which each of flow of liquid, log block design, marbles (taking this as a proxy for the joint component with cards) and covariation were treated as distinct and direct predictors; (b) a partial mediation model, in which the effect of marbles was partially mediated by flow of liquid (i.e. treating awareness of probability as in part supporting spatial-temporal analysis; given the overall predictive strength of flow of liquid relative to marbles, the reverse relationship was conceptually less plausible), and the effect of log block design was partially mediated by covariation; and (c) an extended mediation model, in which the effects of log block design and covariation fed into marbles (i.e. treating awareness of covariation as supporting probability, and through that spatial-temporal analysis). Age and vocabulary were treated as control variables in each case. The last model provided the best fit for the causal total measure.

Further moderation analyses confirmed that there were no interaction effects between log block design or marbles and flow of liquid in predicting causal scores. There was, however, an interaction between flow of liquid and covariation at middle and lower levels of causal scores [for log block design, R-squared=.46, p=.13; for marbles, R-squared=.38, p=.13. However, there was a conditional effect of covariation, R-squared=.42, p=.03, on lower and middle flow of liquid groups, t=4.33, p<.01, but not on the higher group, t=-.70, p=.48. There was no interaction between marbles and covariation (R-squared=.39, p=.12)]. These confirmed that the effects of the measures were independent.

3.5.3 Identification of causal mechanisms

The regression and path analyses illustrated that flow of liquid predicted causal scores consistently, except for inference. Table 16 shows specifically how perfect reconstruction of the spatial-temporal states link to levels of causal explanation. In particular, of those who mentioned mechanism in any task (i.e. explanation scores of 3), 85.7% had a perfect flow of liquid score; of those who never mentioned mechanism, it was 65.1%. The chi-square results for explanation score by perfect versus non-perfect flow of liquid score
were significant for absorption, chi-square=16.549, \( p=0.001 \), and solution, chi-square=9.102, \( p=0.028 \), though not for sinking, chi-square=4.154, \( p=0.245 \); \( df=3 \) for all. These data illustrate at a high level of resolution for the hypothesised link between spatial-temporal analysis and inference of mechanism.

Table 16. Percentages of children obtaining perfect scores on the flow of liquid task at each level of explanation response

<table>
<thead>
<tr>
<th>Explanation score</th>
<th>Causal task 3 (mechanism)</th>
<th>2 (variable)</th>
<th>1 (factor)</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinking</td>
<td>77.8</td>
<td>72.5</td>
<td>60.0</td>
<td>46.0</td>
</tr>
<tr>
<td>Absorption</td>
<td>86.7</td>
<td>76.6</td>
<td>63.6</td>
<td>29.4</td>
</tr>
<tr>
<td>Solution</td>
<td>100.0</td>
<td>79.2</td>
<td>69.2</td>
<td>53.7</td>
</tr>
</tbody>
</table>

3.6 Discussion

3.6.1 The development of causal reasoning about continuous processes

Differences between the three causal tasks. Children’s reports improved with age for all three causal tasks. However, there were developmental differences as well, with sinking easiest, and solution most difficult. Whether dissolving process is more difficult in general for children, or the task differences here reflect issues with the specific examples of the processes chosen remains open. These three processes were chosen in part for their obvious contrast in relative speed, on the view that faster processes, with lower demands on sustained attention, might be easier for children, and the outcomes were consistent with this. It took 1-3 seconds for the stone and berry to sink, 3-5 seconds for the water to rise through the tissue/blotting papers, but 40-90 seconds for the smaller and larger grains of salt to visibly begin to dissolve. However, there are other possible accounts. Solution was also harder to see because of the smaller scale at which it occurred; it involves a less accessible, sub-microscopic mechanism (Liu & Lesniak, 2006); and although children encounter many instances of dissolving in everyday life (cocoa, sweets, soap), these may be less frequent than instances of sinking/absorption.
Future research could directly test whether the speed of the process influences children’s ability to track it during observation – consistent with Maurice-Naville and Montangero’s (1992) reports of elementary school children’s difficulties grasping the effects of slowly progressing forest disease – and whether longer timeframes demand more concentration – consistent with Rieber’s (1991) finding that fourth graders only benefited from an animated demonstration of Newton’s second law when this was presented in short sequences.

**Differences between components of causal tasks.** Results demonstrated a clear progression with age on all causal components. The different components differed widely in difficulty, however, with description ahead of prior knowledge, and explanation trailing far behind. Just under 80% of children across all ages and tasks made and reported completely accurate observations; even for the most difficult dissolving process, 31% of the youngest children achieved the top score for description. For prior knowledge, nearly 58% of children got the highest scores overall. The lack of progress between Y3 and Y5 on these two components therefore seems largely attributable to good levels of performance having already been achieved by 8 year olds. Since children did well on description, this indicates that they had the language for observation, and if prediction lagged behind, it seems more likely to be due to lack of knowledge than lack of words.

Causal explanation scores trailed by long way, with only about 9% reaching top scores across the three tasks and all age groups, and mechanism responses were comparatively rare. Only 5% of children, the majority in the oldest age group, identified any underlying causal mechanism – although once that insight was achieved, it appeared to be quickly extended to multiple processes. Children’s explanations were typically limited to abstracting causal variables from the observed objects, and the youngest age group was commonly unable to explicitly relate causal factors to the observed speed differences, i.e. to actually treat these as variables. Variable abstraction implies increased selective attention to the perceptual input across observed instances of processes, but identification of mechanisms to account for these variables imposes substantially greater inferential demands. For causal explanation, therefore, the lack of progress between Y3 and Y5 is
attributable to children finding it difficult to move on from abstraction of variables to identification of mechanisms.

One can suggest that children’s mechanism responses might have been the product of instruction (at school or at home) rather than inference. However, note that four points indicate this is unlikely: (1) the relatively weak relationship of prior knowledge to explanation, compared to that between description and explanation, suggests responses were constructed largely from observation; (2) although they were infrequent, mechanism responses were given by some younger participants, who would not yet have had more technical instruction in concepts like density, porosity and solubility; (3) conversely, despite the fact that the topics were within the curriculum content, mechanism responses were mostly given by a minority of Y5 children, and there were clear differences between sinking, absorption and solution, again suggesting that observing the contrast phenomena was more important – the difficulties in perceiving the solution events may be a more plausible source of these variations; and (4) provision of mechanism responses was influenced by nonverbal ability, while effects of instruction seem more likely to be mediated by verbal ability.

However, if observation was central to children’s performance, there is plainly a long way from good descriptive ability to higher-level inferential responses. The striking lag in the explanation component may be because explanation of mechanisms places higher demands on children’s language. We know from the pre-school studies involving distinct causal systems that in principle children of this age have no limitation in thinking about mechanism (Buchanan & Sobel, 2011; Legare, Gelman, Wellman, 2010; Schlottmann, 1999; Shultz, 1982). Further investigation is therefore needed to establish how far prompting children to give thorough explanations mitigated the verbal demands of these three tasks (cf. the replication study in chapter four). However, as discussed below, there is good reason to think that language skills per se were not the issue. It seems more likely that children find continuous processes hard to analyse at this level. Comparisons with adults using our paradigm are needed, since mechanism reports may be sparse even among them; no research has examined this to date.
3.6.2 Do verbal and nonverbal abilities predict children’s reasoning about continuous causal processes?

Verbal ability, as indexed by vocabulary, did not predict performance on any of the three causal components, or mechanism responses, arguably the most verbally demanding level of inference. This might potentially reflect the broadly homogenous nature of the sample, with relatively high socioeconomic background – except that vocabulary scores were normally distributed, and there was no indication that variance in verbal ability was limited enough to prevent it having predictive power. Note that the verbal task used in this study was generic, with a focus on everyday language. During testing, it was obvious that children who reported causal mechanisms tended to use more abstract/scientific vocabulary. Such language may be more specifically related to causal reasoning, and could therefore be more discriminating. Dündar-Coecke and Tolmie (2019a) present preliminary data suggesting that the relationship is real on behalf of reasoning about causal processes, though no investigation has focused on this previously, again, suggesting a gap in the literature.

Although it is not entirely clear whether grasp of causal processes is linked to verbal competence, on these data its impact appears to be limited. One possibility is that nonverbal impressions from previously and currently observed phenomena are more critical than explicit concepts, and that these impressions are combined and then translated as far as possible into the verbal domain (see Tolmie & Dündar-Coecke, in press, for a discussion) – meaning that a certain level of verbal ability is required to get this off the ground, in line with the mediation effect in the path model for mechanism. Scientific vocabulary may assist further with this process of translation.

In contrast to verbal ability, nonverbal ability as indexed by block design uniquely predicted causal total, prior knowledge and mechanism. However, the logarithmic relationship of block scores to causal scores reflects a steeper gradient of relationship at lower levels of performance on the latter, indicating that this is where generic nonverbal ability made most difference. This is consistent with the lower explained variance in the regression model for the highest level of explanation response, mechanism, suggesting
that at this level in particular, some additional nonverbal factor might be a key element in children’s cognition of continuous causal processes.

The influence of nonverbal ability on reasoning about continuous causal processes may seem surprising at first. However, to go beyond the observation of a continuous process, as is necessary for thinking about causal variables and connecting mechanisms, requires mental imagery, combining both visible (objects) and intervening invisible (e.g. density, buoyancy) elements. The role of nonverbal ability in this kind of thinking has been explored in another study (Dündar-Coecke and Tolmie, 2019a), which also found that higher levels of reasoning about causal effects do not rely primarily on verbal but on nonverbal competences even across a larger and mixed sample (N=231), suggesting that the finding here is robust. These competences may further include detailed analysis of the spatial-temporal characteristics of processes: the ability to extract key dimensions of information from object states that change over time, to conceive of the sequence of dynamic transformations that underlie such observed change, and to project these transformations onto past, present, and future experiences. This segmentation of a continuous observation into meaningful steps is potentially the additional element suggested above. The following subheading will evaluate this view further.

3.6.3 The development of spatial-temporal analysis

The novel tasks employed to assess children’s spatial-temporal analysis revealed reasonable performance and a regular developmental progression. This study considered an analytical (liquid flow) and a perceptual (speed) spatial-temporal task, with very different processing characteristics. Flow of liquid involved a coordinated spatial-temporal transformation in a single trial, slowed down and segmented. The flow at successive states prompted children for accurate observation and depiction of each. This form of analysis -sequential states- seemed to be helpful, enabling children to extract the principle that as the top level decreases over time, the bottom level increases, and to reconstruct the sequence after the link between top and bottom was broken.

As Piaget (1969/2006) argued, children’s duration judgments are influenced by liquid flow during the intervals. One way to judge this is employing some high level cognitive
processes such as working memory and attention. However, during testing, children with perfect scores seemed to tend to solve the task quickly from logical correspondences: forming a decreasing/increasing sequence for the top/bottom half, then matching the bottom/top; or placing the initial and final pairs, then filling in intermediate states. Children who made errors tended to solve the tasks from memory, and were much slower and unsystematic at the final stage. In this way, flow of liquid involved an analytic response, was easy for most children who understood the logical correspondences, such as how the drawings related to their observations.

Older children were more accurate at organising their representations and locating them in sequential order. Whether this necessitated qualitative (cognitive) or quantitative (conceptual) changes was not clear. Looking at the speed task results, in contrast, which did not require equivalent support or analytical inference; involved 17 trials, each with a brief spatial transformation unfolding over 4-6 seconds; children responded rapidly, within a few seconds. Thus, the two tasks required different processing. The average difficulty level of the speed task - perceptual/attentional- was clearly higher than the analytical liquid flow task, however, which might be attributable to various contrasting features. The three bunny objects in the speed task moved entirely independently, while the states of the top and bottom flask were linked. Thus accurate responses in the speed task depend on processing the successive spatial and temporal states for each object. While both logical and causal constraints may support correct responding in the flow of liquid task, speed relied substantially on attentional and short term memory resources then, in particular majority of young children failed.

The comparison between the liquid flow and speed task responses may provide an answer to the above question (‘whether organizing representations and locating them in sequential order necessitated cognitive or conceptual changes’): while even young children were more accurate at organising their representations and locating them by relying on logical correspondences (i.e. flow of liquid), they necessitated qualitative (cognitive) changes underpinning memory and attentional resources. This is parallel with the cognitive account, suggesting that age related changes in temporal cognition highly linked to changes in attentional processes (see e.g. Driot-Volet, 2011; McCormack, 2015
for a review). To investigate these points further, a subsequent study (N=52) was conducted to compare the nursery (n=17) and reception (n=19) age range responses with adults’ (n=16) on the three different versions of the speed task (original speed; less intensive version with two bunnies but distance widened, and with the actual materials). Initial analysis indicated that children did not exhibit difficulties with the same task when the actual materials were used (i.e. three windup toys). However, they plainly struggled with the computerized versions of the same task, unlike adults who did not exhibit any difficulties with the versions of the task. This study showed that children’s ability to extract spatial-temporal information differed in virtual versus real environments. The computerized version seemed to exclude one of the dimensions, depth, that young children could not compensate. Participants from all age groups did relatively well with the actual task, though with a large variance in responses on the virtual versions. In particular young children found comparing the speed of the three objects on a computerized display most difficult. Some reception children expressed that when the distance between the bunnies was larger, following the speed of the two bunnies was harder. For some children, following the motion of two objects was challenging when the distance was widened between them then (see Appendix 1, for further discussions).

In this subsequent study, when the speed task was administrated with the actual materials, some reception children were capable of distinguishing between distance and duration. This finding has contradicted with Piaget’s (1969/2006; also Piaget & Inhelder’s, 1971) view, who found that young children judged the duration with the distance taken, and argued that this was the result of their confusion between spatial and temporal dimensions, as they assumed longer distance equal to longer duration. He viewed that young children do not reliably distinguish more complex spatial-temporal characteristics such as velocity until about age nine. As discussed in the literature view, other studies found young children demonstrate early implicit knowledge of time, speed, and duration when the tasks involved more practical elements with which children have direct contact in their life (see e.g. Bullock, Gelman, & Baillargeon, 1982; Wilkening, 1981). According to Siegler and Richards (1979), distance-travelled cue in young children’s judgment is affected by spatial characteristics, what develops by age is that once children get older they rely less on spatial elements in their temporal judgments. To test this, later Arlin (1989) isolated the spatial cue by lifting objects of different weights for fixed
durations, he found that spatial cues did not affect duration judgments alone, unlike older ones young children’s judgments were affected by other types of manipulations.

In consistent with Arlin’s argument, the virtual speed tasks (discussed in Appendix 1) required the ability to deal with the type of spatial-temporal manipulations. Children indeed seemed to highly rely on cognitive resources such as attention and memory. This also echoes with the cognitive account (see Driot-Volet, 2011) in a way that age related changes in spatial-temporal cognition linked to changes in memory and attentional processes. However, this was not the only factor, as research indicates that sequencing the past may inherently be easier than predicting the future (cf. Burns et al., 2018; Friedman, 2003; McCormack & Hanley, 2011). This was probably the case for the virtual version, but not for the actual task. The reason might be that in the virtual tasks, (1) compensating for the missing dimension is challenging in particular for young children. (2) The participants needed to calibrate of the horizontal distance for three objects in an asymmetrical setting. Comparing children’s versus adults’ responses to the actual versus computerized speed, young children did not seem capable to deal with asymmetrical manipulations at virtual space, and (3) they seem to be fragile in their approximations. Probably for that reason they tended to follow the moving objects by hands and/or requested to repeat some of the trials. On the other hand, children did not find difficult to predict future relations with the actual materials. This favors the proposal that spatial-temporal characteristics of the presentations do matter when their future outcomes need to be predicted in virtual versus actual environments. Extracting spatial-temporal information from virtual settings might be much harder probably for various reasons. This study shows that (4) calibrating virtual units is more effortful than their real-world properties, therefore (5) they are computationally more intensive, and (6) the elements are considered as immaterial, fast, and hypothetical by the reasoners.

The results resonate with Reiner’s (2018) findings highlighting that it is very likely that virtual tasks require the involvement of other higher mental domains to support performance. Now, we have a slight idea about why does this occur: processing virtual spatial-temporal properties is computationally intensive. And, evolutionary our ability to
extract spatial temporal information from real environment seems to be more advanced. This result, of course, merits further investigations.

3.6.4 What does predict children’s reasoning of continuous causal processes?

In this study, the spatial-temporal flow of liquid task distinguished between lower and higher levels of overall causal performance across the age range, and was the only consistent predictor of higher levels of causal explanation. It was the only measure consistently sharing little variance with the other predictors, including the spatial measures, confirming its distinctness from these. Of two spatial tasks, only one, monkey rotation, predicted one of the components, description, independently, and its influence was restricted to the youngest age group. The other spatial measure, paper folding, was not related to any indices. This thesis acknowledges the fact that spatial ability can take various forms, the most robust and studied kinds are either mental rotation or mental transformation, but their role in causal reasoning about continuous processes seems limited, and appears to overlap in particular with verbal ability. This finding is contrary to Paivio’s (1971, 2013) proposal, suggesting a dual route between verbal and spatial reasoning. The lack of predictive power might seem to originate from the fact that most spatial tasks rely on two-dimensional representations and do not carry temporal information. Missing the temporal element may cause them to be encoded in a discrete fashion unlike on-line fluid spatial-temporal representations. Kosslyn (1980; see also Pylyshyn, 2003; Kosslyn et al., 1983) also argues that spatial representations are quasi-pictorial. These points may explain the predictive power of the spatial tasks for the descriptive ability. However, the discrete nature of these tasks seems to be restrictive when further feedback is needed for the past, current and future representations.

Hegarty uses mechanical examples that people mentally need to rotate the objects and imagine/simulate the outcomes. For example, in her 2004 study she demonstrates a sequence of gears connected to each other in which participants need to rotate them by following the clockwise direction. One can suggest that in this form of spatial demonstration some sort of causality is involved in mental rotation (e.g. Lagnado, 2011, for a review). For instance, in Hegarty’s study, the initial clockwise motion (cause) leads physical material to be rotated (final effect). The present study acknowledges this view,
and considers that sometimes it is possible to add temporal dimension into static pictorial tasks. However, even in this kind of analysis causal factors are piecemeal, organised discontinuously, they do not operate on a holistic fashion that allows feedback. Moreover, including the temporal dimension qualitatively turns them into a dynamic spatial-temporal task, since adding temporal information changes representations regarding the patterns and connections between them. There are of course other kinds of spatial representations. For instance, in a causal map, the effect has to follow its cause (precedence) (see Gopnik et al., 2004). Coming from topological foundations, in a diagrammatic language, a path is a process by means of time. On the other hand, various combinations of spatial-temporal properties can be displayed via two-dimensional settings, such as movies. The nature of the mental effort changes when temporal information is involved. But apparently, spatial tasks do not inform causal processes as soon as they rely on solely mental rotation or transformations and lack crucial temporal components such as contiguity/precedence, succession, continuity, or transformation of the movement. Once they involve one of these temporal elements, they cannot be defined as ‘spatial’ though.

Nonverbal ability, in contrast, as indexed by block design, did predict reasoning about causal processes independently of spatial-temporal ability. However, its influence was consistently restricted to discriminating at lower levels of causal performance (as the trends were consistently logarithmic). This may involve a bootstrapping effect, probably the detection and analysis of spatial elements. But patterns superseded by spatial-temporal analysis and this seemed to be the driver of later causal reasoning across development (cf. the identification of factors and variables). Verbal ability did not predict causal performance in this study, which had a quite homogenous sample with relatively high socioeconomic background (nonverbal ability was more discriminating in this case). This point will also be investigated further in the replication study.
4.1 The rationale of the replication study

The previous study had four key findings, as broadly discussed under the relevant subheadings. First of all, the results confirmed developmental trends in the ability to analyse various kinds of information, with clear increases across the age groups. However, children exhibited uneven profiles on some of the measures. Results clarified that (i) nonverbal, (ii) statistical, and (iii) solution measures showed later growth than the verbal, spatial, and one of the spatial-temporal measures (i.e. children approaching at ceiling by Y5 on flow of liquid), and thus suggesting a late development in the competencies measured by those three tasks. The sample spanned the normal range, but with above average means for in particular verbal, and nonverbal ability. Comparing with WASI age equivalent scores, the mean age equivalent for verbal Year 1 was 8 years, 9 months; for nonverbal Year 1 was 7,2; for Year 3: 11,9 and 8,8; and for Year 5, 14,11 and 11,2 respectively. This showed that the sample was high achieving when international average scores were considered.

Regarding the chemistry measure, solution, there is no readily available study to compare the outcomes, as discussed in the first chapter. However, this finding –late developmental trend in solution- favours Adams and Griffard’s (2001) analysis of children’s conceptions for biological and physical indices (cf. Chapter 1) in the sense that topics in biology and physics may be more interconnected, and in particular biological concepts tend to be more language sensitive. Given that verbal measure scores of the sample were above average and children did better in sinking, and relatively well with absorption, chemical phenomena may be more challenging for children due to several reasons. The difficulty may not originate from the lack of sensitivity to invisible particles in dissolving. As Rosen and Rozin’s (1993) study indicated that even preschoolers could acknowledge the presence of an entity even in the absence of a visual residue. Moreover, all three
phenomena involved invisible processes (e.g. gravity, upthrust in sinking, capillary action in absorption), but children in particular struggled with the solution. This brings to mind Wellman, Hickling, Schult’s (1997) argument underlining that children may have advance explanatory reasoning systems for biological and physical phenomena. However, these studies do not explain why there are different developmental trend for each causal task. Although various possibilities were discussed (i.e. the characteristics of the tasks, domain differences, speed of the processes), this point merits another investigation with a broader sample, which includes children from different socioeconomic backgrounds to elaborate on whether differences in children’s explanations are consistent across the three tasks, as well as their verbal, nonverbal, and statistical measures.

Another key finding was that the ability to analyse spatial-temporal transformations improved with age. The developmental trends for flow of liquid suggested that children were better able to reconstruct a sequence of spatial states across time. Speed results indicated that the ability to register online rapid movements and extrapolate future positions from computerized displays were exhibited later in development, and therefore appeared to be more difficult. However, the ability to reconstruct a spatial-temporal sequence was a crucial predictor of children’s causal performance as a whole; this was true also when the same analyses were run with prediction, description and explanation elements separately. Regarding the components, the predictive power of flow of liquid measure was limited to prior knowledge and descriptive level responses. For explanation the patterns were more complex, with no predictors in the final model. However further statistical analysis indicated that for mechanism level responses, flow of liquid was the most robust predictor. The other spatial-temporal measure, speed, did very little. For now, it remains unclear as to which forms of spatial-temporal analysis are important to children’s reasoning about natural causal processes beyond analytical segmentation. Whether the predictive part is a narrower or different property of the flow of liquid task; or if greater exposure to instances of each process would change the picture.

To investigate these further, in this study, the flow of liquid task was retained. However, the water level changed in six steps rather than five, to enhance its discriminating power among older children who were typically at ceiling on the five-step version. The speed
task was replaced by a set of tasks elaborating judgments on one of the following: distance, time, or velocity (DTV) from information about the other two of these dimensions, based on Wilkening’s (1981) study. The DTV instead required mental simulation of movement against a time count. This aimed to elaborate on whether it was spatial-temporal analysis involving mental construction of a sequence of change that was predictive of causal performance.

Wilkening (1981) used velocity, time, and distance as a set of interrelated dimensions. Velocity was defined as a relation between time and distance; distance was conceptualized by time. In Piaget’s work, those three concepts were investigated exclusively, and coordinated metrically. A child needed to coordinate these concepts via deductive operations, which appeared very late towards the end of childhood. Wilkening’s criticism to Piagetian approach is that with this methodology it is not clear if children understood about the relations among these variables. Therefore it remains unclear whether the child understands the main conceptual questions. For instance, in Piaget’s distance-time travelled example, where children were asked which of the two trains travelled the greater distance, Wilkening proposes that Piaget asked the questions in such a way that time and speed information served to distract the child’s attention from the distance cue. The interrelationships between the concepts were ignored. To avoid these problems, Wilkening’s methodology represents the concepts in different ways: information about two dimensions is presented in separate events, and the child needs to infer the third dimension from these cues. Wilkening’s methodology therefore proposes to integrate stimulus dimensions in a more age appropriate way, as even young children seem to integrate velocity, time, and distance information. For that reason, the present chapter will employ his integrated dimensions approach, where each dimension is multiplied and subtracted in a holistic process.

Another key finding was that in that sample only nonverbal and spatial-temporal measures were stable predictors of children’s reasoning of causal processes. The classical and frequentist probability tasks were found to be closely relevant to each other, but none of the statistical measures could survive in the regressions. This might be partly attributable to the limited number of steps involved in the task affecting their sensitivity,
or alternatively the existence of much stronger predictors in the regression models may affect their predictive power (flow of liquid and block design largely subsumed the explained variances of marbles, and both probability tasks interacted with flow of liquid at lower ability level). The replication study aimed at investigating this matter with more refined statistical measures. The marbles task was retained, due to its potential captured by the path analysis. The covariation task was modified. While in supportive circumstances even young children may be able to assess covariation from all four cells of the contingency table, if tasks have any complexity elementary school aged children fall back on strategies considering only the relative frequency of positive co-occurrence (AB versus AnotB cells). For the present study, a more child friendly non-causal covariation task was devised, in order to assess whether individual differences in covariation assessment predict children’s causal thinking. The link between the variables relied on the statistical information, but without support by potential knowledge of a causal relation. In all other respects, the task was as simple as possible, using a pictorial approach, consistent with the literature and with the other tasks in the battery, to investigate children’s assessment of simple covariation patterns (as shown in Figure 17).

Since the monkey rotation task failed to discriminate older children’s performance, it was replaced by Stefanatos, Buchholz, and Miller’s (1998) task adapted from Shepard and Metzler (1971), which demonstrated good discrimination across the elementary age range performance. It also used physical materials, providing a better parallel to the materials used in the battery. Given the overlaps between paper folding and rotation tasks in the previous study, and its limited predictive power, the paper folding task was dropped.

Beyond these tasks modifications, this chapter extended the methodology and employed a more diverse sample to test the same hypothesis and the research questions for greater generality. Regarding the methodology, the causal tasks required more observations, and corresponded to a more realistic scientific procedure, with the focus again on contrasting speed of effect within sinking, absorption and solution. The scientific procedure followed this sequence: observation, prediction, testing, justification, and explanation. The observation and prediction stages were as before, but there was no initial prediction from prior knowledge. Instead, prediction for three further objects followed initial
observations, so could be informed by these as well as prior knowledge. Children were also asked to justify why they thought their predictions were correct. They then tested their predictions and finally gave explanations of what they had observed. By following this sequence, children had a more active role in the procedure, and opportunity to think about each process across five rather than two examples before concluding causal inferences, but still without intervening. This approach aimed to elaborate on whether increased exposure promoted better causal performance in this context. Pilot testing (with four children) indicated that children gave more extensive explanations for these five-item comparisons. This allowed responses to move from ‘simple identification of causal variables’ to differentiate ‘the joint operation’ of variables. In other words, long before children mentioned about mechanism, they needed to coordinate further information available during observation, prediction, and justification stages. The scoring system for explanation therefore captured this difference and examined whether it constituted a further interim step in children’s progress towards awareness of mechanism.

In this methodology, children were asked to watch the sinking/absorption/solution of two objects; they were requested to describe their observations ‘what happened’, and then were presented with three more objects. In sinking, these objects were a marble, a play dough ball, and a cherry tomato; in absorption the further object were a piece of cardboard, fabric, and foam; and in solution they were three kinds of sugar (muscovado, demerara, and powder). In all these, one dimension (volume) varied at a time, while other variables were kept constant, such as either colour or shape. A similar approach was used by Piaget and Inhelder (1942/1974). They paired a large cork with a smaller but heavier stone. In Smith at al.’s (1985) experimentation a similar approach was used. The authors named their approach ‘critical pair’ errors. A critical pair was, for instance, a small, light piece of steel against to a large heavy piece of aluminium. In both studies, children misjudged that the biggest object would be heavier than the smaller one. Kloos and Somerville (2001) also employed a similar approach to study children’s misconceptions and conceptual change (i.e. how volume predicts sinking speed). Children were shown a pair of toy submarines either holding two, three, four, or five units of weight, or they differed in their overall area (ranging from 30 to 90 cm2). The pairs of objects differed in volume, while mass was held constant. In two conditions, some children were received demonstrations and were given the opportunity of interviews to form expectations about
how size and weight separately relate to sinking speed. Others did not receive demonstrations. Children in both conditions were required to learn how volume affects sinking speed in a way that was opposite to their original beliefs. The authors found that children, those who received demonstrations and were interviewed about the effect of size on sinking, were able to refine their belief about the effect of weight, suggesting that a relatively short demonstration helped children to refine their beliefs.

This conclusion was tested by Kloos and Van Orden (2005). Considering the literature showing that young children consistently attribute sinking speed to be dependent on differences either in mass alone or in volume, the authors created a more complex laboratory condition. They attempted to teach preschoolers that neither mass nor volume is predictive for sinking speed alone, and then tested whether beliefs can change when the correct dimension is never pointed out specifically (as in Kloos and Somerville’s 2001 study). The authors found that young children can alter their mistaken beliefs. However, when mass is pitted against volume they can still demonstrate a mass bias, possibly emerging out of their previous experiences with the force (e.g. falling objects, heavier objects require greater lifting). This also indicates that mass may be a more salient dimension than volume.

The aim of this chapter is to take one step further and investigate whether children’s predictions improve or worsen when both volume and weight are pitted against density (i.e. presenting three further objects varying in size and weight). The assumption is that the possibility of detecting a meaningful pattern in the new observation would be possible when prior demonstrations are made. This may lead to an expectation that children’s prediction scores may improve when size and weight separately relate to the sinking speed (i.e. smaller object sank faster than the larger ones). The question is therefore whether dealing with the new evidence can help children to interpret causal processes.

Scientists calculate various objects’ density/porousness/compactness by using apparatus, even taking into account the temperature and what the object is made of. Although children were asked to weigh each object by hand and feel the texture, strength, or structure, they did not use extra apparatus. However, the volume and the weight of the
materials in sinking, the porousness and the type of the material in absorption, and the size and surface area of the three kinds of sugar in dissolving were made salient to trigger children’s thinking about various variables distinctly. Moreover, children were able to see whether their predictions were correct, and even justify their choices after this observation, allowing them to evaluate/alter their views afterwards. These stages were aimed at providing the children with more experience and getting them acquainted with the phenomena before the explanation across three phenomena.

4.2 Method

4.2.1 Participants

Participants were 124 children from three primary schools in Oxford (36 from Y1, mean age = 5 years, 11 months, sd = 3.8 months; 45 from Y3, mean age = 7 years, 11 months, sd = 3.6 months; 43 from Y5, mean age = 9 years, 9 months, sd = 5.1 months). Of these, 65 (52.4%) had a monolingual (English) home environment, and 59 (47.6%) came from bilingual/multilingual homes, exhibiting similar variation to previous study in Chapter three. However, the socioeconomic background was more broadly representative, normally distributed, with 20 (16.1%) of parents being unemployed, 32 (25.8%) manual workers, 48 (38.7%) self-employed/non-manual workers, and 24 (19.4%) professionals. In total, 35 parents (28.2%) had only GCSE qualifications; 21 (16.9%) had A levels; 38 (30.6%) had an undergraduate degree; 22 (17.7%) had a postgraduate/professional qualification; and 8 (6.5%) had doctoral degrees.

4.2.2 Design, Materials and Procedure

Design, age groups and task order were as in previous study. Tasks were given children in the same fixed order within a single one-to-one sessions: measures of verbal and nonverbal ability were assessed as before, three causal tasks, two spatial-temporal tasks, one spatial task, and two statistical thinking tasks. An average test session length was 41 minutes (min = 28, max = 56). Children were told that they could take a break whenever they needed, but only three of them expressed the need for a break across whole sample. This can be an indicator for the length of the testing session was being appropriate.
4.2.2.1 Causal tasks
The causal tasks again focused in turn on contrasting instances of sinking, absorption and solution, but in this case children first observed and described two instances before being asked to predict, with justification, outcomes for a further three, and then make inferences about the influences at work across all five instances. Three types of measure were computed from these tasks. (1) Composite scores: totals for sinking, absorption, and solution. (2) Components: totals for accurate description, prediction, justification, and explanation separately (basic factors, operative variables, relationships between variables, and mechanisms); and (3) a total score for causal performance across all indices, 0-33, alpha = .724. Interest again centered on the overall score for causal performance and that for inference.

Each causal task had five stages: children (1) inspected the two contrasting materials from Study 1 (except a grape replaced the blueberry), watched them sink/absorb/dissolve in water, and described what they saw; (2) inspected three further contrasting materials (Figure 15) that sank/absorbed/dissolved at different rates, and predicted which would sink/allow water to rise/disappear fastest, next fastest and slowest; (3) justified why they thought their predicted order was correct; (4) tested their predictions with the aid of the experimenter, to evaluate whether they were correct, and finally (5) explained why they thought things had happened in the way they had seen across the various observations. At inference stage children were encouraged to consider the initial items (e.g. stone, grape) to assure the consistency with the previous study.

For sinking children saw a stone and a grape of similar size and colour but different densities, which sank at different rates in a half meter tall large transparent jar of still water. The five-stage design implemented by the following scripts:

Description of observation: “I've got these two objects here; a stone and a grape. I’m going to drop them in the water. Please watch carefully.” Did you notice anything, what?

Prediction: “Now, I will show you these three objects, marble, playdough ball, cherry tomato. Can you rank their sinking order/rate, which one will sink fastest and which one will sink slowest? Which one will be in the middle?”

Judgment: “Why did you rank/order them in that way?”
**Testing:** “Could you test these to see if your prediction is correct?” “Was your prediction correct?”

**Explanation:** “Why do you think things happened that way?” “Do you think there might be another reason for that?”

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinking</td>
<td>(marble, plasticine, tomato)</td>
</tr>
<tr>
<td>Absorption</td>
<td>(fabric, cardboard, polystyrene)</td>
</tr>
<tr>
<td>Solution</td>
<td>(caster, muscavado, demerara sugar)</td>
</tr>
</tbody>
</table>

Figure 15. Materials in the correct order of prediction for the causal measures

For *absorption*, children saw water rising from a petrie dish through the strips of tissue and blotting paper of the same length/width, the water rising faster through the more open structure of the tissue. Similar to sinking protocol, the five-stage design implemented:

**Description of observation:** “I've got these two strips of paper here; a tissue paper and a blotting paper. I’m going to dip them in the water. Please watch carefully.” Did you notice anything, what?”

**Prediction:** “Now, I will show you these three objects, a piece of fabric, a piece of cardboard, and a piece of foam. Can you rank them, which one will soak up water fastest
and more and which one will soak up slowest? Which one will be in the middle?"

_Judgment_: “Why did you rank/order them in that way?”

_Testing_: “Could you test these to see if your prediction is correct?” “Was your prediction correct?”

_Explanation_: “Why do you think things happened that way?” “Do you think there might be another reason for that?”

For _solution_, children saw the same small quantities of table and rock salt dissolve in warm water. The small quantity of the salt was assured with two equally very small spoons. The greater surface area to volume of the table salt led to more rapid solution. Similar to sinking and absorption protocols, the five-stage design implemented by the following scripts:

_Description of observation_: “I’ve got two types of salt here; some table salt and some rocky salt. I’m going to put them in the water. Please watch carefully.” Did you notice anything, what?”

_Prediction_: “Now, I will show you these three objects, muscovado sugar, demerara sugar, and caster sugar. Can you rank their disappearing time, which one will disappear fastest, which one will be the slowest, and which one will be in the middle?”

_Judgment_: “Why did you rank/order them in that way?”

_Testing_: “Could you test these to see if your prediction is correct?” “Was your prediction correct?”

_Explanation_: “Why do you think things happened that way?” “Do you think there might be another reason for that?”

Scores for children’s description, prediction, justification, and explanation responses were arrived at in the same way as the previous study (cf. Chapter three). Using the modified system shown in the Table 17, composite measures were computed for each response component as follows:

0-3 for _description_
0-9 for prediction and justification

0-12 for explanation

0-33 for total causal score (alpha = .724).

0-3 for the number of mechanism level responses made by children.

Table 17. Scoring system for causal task

<table>
<thead>
<tr>
<th>Description of observation (0-1)</th>
<th>Sinking</th>
<th>Absorption</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>No observation=0</td>
<td>No observation=0</td>
<td>No observation=0</td>
<td></td>
</tr>
<tr>
<td>Observing different sinking rate=1</td>
<td>Observing different rate of water rising=1</td>
<td>Observing different solution rate=1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction following observation (0-3)</th>
<th>Sinking</th>
<th>Absorption</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any prediction plasticine comes first=0</td>
<td>Any prediction foam comes first=0</td>
<td>Any prediction other than below=0</td>
<td></td>
</tr>
<tr>
<td>Any prediction tomato comes first=1</td>
<td>Fabric-foam-cardboard=1</td>
<td>All same=1</td>
<td></td>
</tr>
<tr>
<td>Marble-tomato-plasticine=2</td>
<td>Cardboard-fabric-foam=2</td>
<td>Caster-demerara-muscavado=2</td>
<td></td>
</tr>
<tr>
<td>Marble-plasticine-tomato=3</td>
<td>Fabric-cardboard-foam=3</td>
<td>Caster-muscavado-demerara=3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Justification of predicted order (0-3)</th>
<th>Sinking</th>
<th>Absorption</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>No/irrelevant explanation=0</td>
<td>No/irrelevant explanation=0</td>
<td>No/irrelevant explanation=0</td>
<td></td>
</tr>
<tr>
<td>Only weight=1a</td>
<td>Only thickness=1a</td>
<td>Only material=1a</td>
<td></td>
</tr>
<tr>
<td>Only size=1b</td>
<td>Only softness=1b</td>
<td>Only size=1b</td>
<td></td>
</tr>
<tr>
<td>Both without coordination=2</td>
<td>Both without coordination=2</td>
<td>Surface area or compactness=2</td>
<td></td>
</tr>
<tr>
<td>Both with coordination=3</td>
<td>Both with</td>
<td>Both coordinated=3</td>
<td></td>
</tr>
<tr>
<td>Explanation (abstraction of causal factor; link to speed difference i.e. variable; coordination of variables; mention of how variable affects speed) (0-4)</td>
<td>No/irrelevant explanation=0</td>
<td>No/irrelevant explanation=0</td>
<td>No/irrelevant explanation=0</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Weight/size without difference=1</td>
<td>Thickness/softness without difference=1</td>
<td>Grain/size etc. without difference=1</td>
<td></td>
</tr>
<tr>
<td>Weight/size with difference=2</td>
<td>Both with difference=2</td>
<td>Grain/size etc. with difference=2</td>
<td></td>
</tr>
<tr>
<td>Density (weight/size coordinated)=3</td>
<td>Structure/holes coordinated)=3</td>
<td>Grain/size etc. coordinated with compactness=3</td>
<td></td>
</tr>
<tr>
<td>Density with mechanism=4</td>
<td>Optimum hole size with mechanism=4</td>
<td>Surface area, compactness and solvent mechanism =4</td>
<td></td>
</tr>
</tbody>
</table>

Example explanation responses illustrating the additional level of scoring for coordination of causal variables are also shown in Table 18.

<table>
<thead>
<tr>
<th>Phenomena</th>
<th>Level 3</th>
<th>Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinking</td>
<td>“The marble is smaller, heavier and harder than the playdough and the dense, the water doesn’t have tomato, and the stone is like the enough strength to make them marble.”</td>
<td>“The marble and the stone are denser than the others that is why they sank faster.”</td>
</tr>
<tr>
<td>Absorption</td>
<td>“The tissue paper and the fabric are lighter and have texture. They have and holes which allows water go room, which allows water to rise into it faster then the blotting paper. The cardboard has holes too.”</td>
<td></td>
</tr>
</tbody>
</table>
but they’re bigger, and it’s harder for the water to spread through them. More compact more difficult for water to rise”

Solution “The table salt is less compact like caster sugar, they are both smaller and softer, and easier to spread around.” “The table salt and caster sugar has less surface area, the water can cover it in no time and break it all down. For the bigger chunks it takes longer for the water to get around it and into it, so it takes longer to dissolve. Some materials (e.g. demerara) have harder walls, compactness.”

4.2.2.2 Measures of spatial-temporal analysis
The flow of liquid task was identical to the previous study, except that the liquid flowed from upper to lower flask in six rather than five steps, giving children six drawings to rearrange at stage two, and 12 at stage three. Scores could therefore range from 0-12.

DTV tasks. Children completed three related distance/time/velocity (DTV) tasks (Wilkening, 1981), displayed in Powerpoint, which required them to estimate of each of distance, time and velocity in that order, by integrating information about the other two dimensions.

For distance judgments, children saw a cartoon cat, mouse, and turtle, and were told these animals differed in speed (cat fastest, turtle slowest, mouse in-between). They then imagined how far each would run in a fixed time, counted out loud by the experimenter, to escape from a barking dog. There was one practice trial of 4 seconds involving just the cat (Figure 16), for a calibration answer (middle of the bridge) was agreed with the child, followed by three test items: all three animals appeared in the same starting positions relative to dog and bridge; the dog barking for 6, 4 and 2 seconds. Children pointed to
where each animal would have run each time. A typical session included the following scripts:

Experimenter for the practice trial: “Now, you will see four animals, a dog, a turtle, a mouse, and a cat. When the dog starts barking, other animals run over the bridge. They stop when the dog stops barking. The cat is the fastest, the mouse is the second fastest, the turtle is the slowest. You will need to show me the point that animal would have reached each time when the dog stops barking. Let’s take a look at these examples. The dog barked 4 seconds, one, two, three, four… Okay, where this cat would have reached? Remember, they’re running over the bridge. Yes, I think you're right, it would get there easily in 4 seconds or No, it’s faster than that, let’s try that again, One, two, three four... It gets there easily.

The experimenter made sure that the child understands that all animals run over the bridge, and they point to the animal’s position and verbally express where they are.

Experimenter for the test trials: “Okay, you’ve got the idea. Now, I will not help you. There will be three different barking time, 2, 4, 6 seconds. Please show me the point that animals would have reached each time when the dog stops barking. Ready to start? Okay great. The dog barked 6 seconds: one, two… six. Where these animals would have reached? How about the cat/mouse/turtle?”

This process is completed for the next two trials. Children received 2 points per item if their answer was fully correct (based on extrapolation from the initial illustration, relative animal speed and time), 1 for partially correct answers (correct distance for some but not all three animals), and 0 for completely incorrect answers. Scores could therefore range from 0-6.

For time judgments, children saw only one animal (cat, or bunny, or turtle respectively) and estimated, by counting out loud themselves, how long this animal would take to run to a fixed point. The first part of the run was animated, with the second half continued
behind a wall. There were two practice trials, bunny and cat (Figure 16), during which children were helped to count appropriately for calibration, followed by three test items in the order turtle, cat, bunny. The cat and bunny in the test items differed in speed from the practice items. A typical session included the following scripts:

Experimenter for the practice trial: “This time, you will see a point here where the animals arrived each time. Look at this bunny. It will take him 5 seconds to run that distance from there to the point, one, two, three, four, five. You will see the half of the run and the bunny will run the other half of the run behind the wall here. You will need to imagine and count how long would it take for bunny to complete that distance. Let’s take a look at these examples. The bunny is running that distance in one, two, three, four, five seconds. Okay, let’s have a look at this cat. Remember to count the running time behind the wall. Look at this cat. How long does it take the cat to run that distance? Yes, I think you’re right, it would get there easily in 5 seconds or No, it’s faster/slower than that.

The experimenter made sure that the child understands that the running time behind the wall needs to be counted.

Experimenter for the test trials: “Now, I will not help you. There will be three different runs. Please tell me how long time each animal run to complete the distance. Ready to start? Okay great. How long does it take this turtle to reach to the point?”

This process is completed for the next two trials. Children received 2 points for per item if they counted to the correct time, 1 if they counted to one second either side, and otherwise 0. Scores therefore ranged from 0-6.
For *velocity*, children had to judge which of seven animals (deer, horse, cat, bunny, mouse, turtle, snail) were fast enough to make it to a house at varying distances in a given period of time, counted out loud by the experimenter. There was one practice trial (Figure 16) in which the animals had 3 seconds to run, and children were helped to the conclusion that just the horse, deer and cat would get to the house mid-screen while ensuring that any judgement of relative speed came from the child themselves. Then they saw three test items, with a) house located mid-screen, 4 seconds, b) house located the close side, 2 seconds, and c) house located the far side, 6 seconds. A typical session included the following conversation:
Experimenter for the practice trial: “Let’s start with this practice, to get you used to doing this. You see the animals are lined up at the bottom on the left, with the house in the middle. They have 3 seconds to run towards it. Remember, they’re only trying to get to the house, not inside it. Let’s start by thinking about each animal. I’ll count the time up to 3 seconds, and you imagine these animals running. One, two, three... Okay, which animal you think will reach the house in three seconds? Yes, I think you’re right, it would get there easily in 3 seconds or No, it’s faster than that, think about three seconds, one, two, three...

The experimenter made sure that the child understands that all animals run to the house, not somewhere else, and they point to the animal’s position and verbally express where they are.

Experimenter for the test trials: “Okay, you’ve got the idea. Each time their distance from the house will be changing. You will need to tell me which animals will reach the house in 2, 4 and 6 seconds. Let’s try this one, and I won’t help you now, I’ll just count the time, you can join me if you want. Here they have 2 seconds. Which ones would have reached the house? How about the rest?”

This structure is repeated for the next two trials. For each item, children were given 2 points if they identified a predetermined set of ‘winners’ (calibrated against the practice trial), 1 if they missed some out or included others, and 0 if they misjudged (e.g. turtle is faster than dear). Scores therefore ranged from 0-6.

A total DTV score (0-18) was also computed across the three tasks, partial correlations controlling for age, $r_{\text{distance/time}}=.417$, $r_{\text{distance/velocity}}=.363$, $r_{\text{time/velocity}}=.180$, $p<.05$ for all. A principal component analysis showed 56.6% shared variance in the first factor, with loadings of .840, .725, .684 for distance, time and velocity respectively. Together with the correlations, this confirmed that each component measured different competences.
4.2.2.3 Measure of spatial ability

In the mental rotation task, children saw two wooden stick figures, Tick and Tock (Figure 16; Stefanatos et al., 1998), and were instructed that they would see photos (colour, pasted on to laminated A5 cards), sometimes of Tick and sometimes of Tock, and had to say which it was. There were three practice trials, during which any errors were corrected. These were followed by 16 test trials (2 x 0°, 4 x 45°, 4 x 90°, 4 x 135°, 2 x 180°, balanced across target and clockwise/anticlockwise rotations), presented in pseudo-random order.

The instruction process was implemented as in Stefanatos et al.’s study. Children were first presented with three practice photographs in front of tick and tock: “I have this set of funny looking twins –tick and tock here- to show you. You don’t need to remember the names, but you need to know that they are mirror images of each other. This means that this part of this goes in this direction and this part of this one goes in that direction. When they face each other, it is like they are looking in the mirror. I will show you another picture of them each time with different positions. Please do not touch them, or rotate your head, just look at them and compare. What do you think that picture is of this one or this one? Let’s do some more practice. What do you think about that one, is it this ne or that one?

Once experimenter convinced that the child understood the task requirements and was ready to give a go, the actual trials began. The actual trials continued the same way with different photographs rotated tick and tock in the ways described above. The score was the number of correct answers (0-16).

4.2.2.4 Measures of statistical ability

The marble task exactly followed the same structure with the same instruction as described in Chapter three. The cards task was not included, as marbles and cards shared the majority of their variances with each other, with marble task was more predictive.

The covariation task utilized decks of cards all showing one of four types of image: either a circle containing a star; or a circle containing a crescent moon; or a square with a star;
or a square with a moon. Attention focused throughout on the degree of co-occurrence between stars and circles, with the squares as distractors. The child saw four decks in turn consisting of eight cards, four cards showing circles and four squares. In the first deck (Figure 17), three of the circles contained stars and one a moon; co-occurrence between stars and circles was therefore 75%. In the second deck, co-occurrence was 50%, in the third it was 100%, and in the fourth it was 0%. In each case, the cards were laid out face up before the child in random order, and they were asked to say from what they saw in front of them how often stars and circles went together. Children had to evaluate whether a particular surround shape (e.g., circle or square) “goes with” particular figure inside (e.g. star or moon).

Figure 17. One of the four trials of the covariation task employed in Study 2, displaying the first deck with 75% co-occurrence between starts and circles

The task was presented to each child with the following instruction:

Experimenter: “Now, you will see decks of cards. Each deck will contain eight cards like this. If you look carefully, there are squares and circles, either goes with the star or a crescent moon here and here…You will need to tell me how often a circle goes together with a star? Yes, that’s right, thanks / no look carefully you need to focus on the circles and stars/moons. Okay. How often a square goes with a moon? Yes, that’s right, thanks / no look carefully you need to focus on the squares and moons/stars. Okay. You can tell
me how likely go together or just by ticking on this line starting from ‘never go together’ here to ‘always go together’ here. Where would you tick on the line?

Once experimenter convinced that the child understood the task requirements and was ready to give a go, the actual trials began. The actual trials continued the same way with different decks utilizing the same shapes (circle and square) and the figures (star and crescent moon). The focus of this task was on whether children could extract this relation in a number of problems differing in the relative frequency of co-occurrence (0%, 25%, 75% or 100% co-occurrence). Correct answers on both response elements (allowance was made for some lack of exact precision) were marked as 2 points for each deck, and partially correct (only qualitative or quantitative element correct) as 1. Scores were therefore ranged between 0 and 8.

4.3 Results

Analyses utilised data from 124 participants. The results are presented in similar sequence to chapter three. Similar to the previous study, two-sided tests were applied where relevant, by using a significance level of .05. The observed power for regression analyses was 0.97. One-way-ANOVAs, zero-order and partial correlations, frequency distributions, skewness, trend analyses were considered. Hierarchical regression, maximum likelihood, and mediation-moderation analyses were reported in the same sequence.

4.3.1 Developmental trajectories

4.3.1.1 Causal tasks

Composite scores. Similar to the previous study, performance was best on sinking, but this time solution was slightly easier than absorption. A two-way mixed ANOVA (task within-subject, age between-subjects) found a significant main effect of task F(2,120)=17.915, p<.001, partial eta-squared=0.230 using a Bonferroni comparison. There was also a main effect of age group, F(2, 121)=6.192, p=.003, partial eta-squared=0.093, with scores for Y1 and Y3 significantly lower that Y5, but no difference between Y1 and
Y3, with no task by age interaction, reflecting greater growth between Y3 and Y5 for all tasks.

Components and causal total scores. Table 19 shows the means for each age group on the causal components and causal total scores. Relative to the scales employed, performance was highest on description and prediction, as in previous chapter. However, performance on explanation lagged behind more than the previous study. Mechanism responses also decreased, with only 15.3% of children giving mechanism level response on one or more tasks here, against 19.6% previously. Most children’s explanation responses focused on relevant variables (scores of 2), some of them explicitly coordinated the variables (scores of 3) – 53% versus 18.3% on average across the three causal tasks. Justification responses were poorer, with the majority restricted simply to identification of a single key variable (72.6% on average).

Table 19. Mean score (sd) by age group on total score (max=33), description (max=3), prediction, justification (max=9), explanation (max=12) and mechanism (max=3)

<table>
<thead>
<tr>
<th></th>
<th>Y1</th>
<th>Y3</th>
<th>Y5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal total</td>
<td>15.80 (5.21)</td>
<td>16.77 (5.12)</td>
<td>19.27 (3.23)</td>
</tr>
<tr>
<td>Description</td>
<td>2.22 (0.83)</td>
<td>2.53 (0.59)</td>
<td>2.74 (0.44)</td>
</tr>
<tr>
<td>Prediction</td>
<td>5.97 (1.93)</td>
<td>5.69 (1.68)</td>
<td>6.42 (1.64)</td>
</tr>
<tr>
<td>Justification</td>
<td>2.58 (1.25)</td>
<td>2.87 (1.29)</td>
<td>3.49 (1.37)</td>
</tr>
<tr>
<td>Explanation</td>
<td>5.03 (2.36)</td>
<td>5.69 (2.56)</td>
<td>6.63 (1.75)</td>
</tr>
<tr>
<td>Mechanism</td>
<td>0.06 (0.23)</td>
<td>0.16 (0.42)</td>
<td>0.28 (0.50)</td>
</tr>
</tbody>
</table>

One-way ANOVAs showed age-related progression for total causal scores, F(2,121)=6.192, p=.003, partial eta squared=.093; and on each component except prediction: for description, F(2,121)=6.811, p=.002, partial eta squared=.101; for justification, F(2,121)=5.065, p=.008, partial eta squared=.077; for explanation, F(2,121)=5.096, p=.008, partial eta squared=.078. In each case, there were significant differences between Y1 and Y5, with Y3 intermediate. There was a marginally
significant age effect on mechanism, with $F(2,121)=2.936$, $p=.057$, partial eta squared=.046.

In general, effects were smaller than the previous study, and scores were lower, comparatively, tending to cluster around the mid-range of the scale; except for description, they were normally distributed. Growth also took place later than the previous study, suggesting that either the tasks were more difficult for children, or the sample had lower ability, or both. Verbal and nonverbal responses were notably lower than the previous study. The lack of age effect for prediction did not seem to be attributable to children reaching ceiling early, but to performance mostly remained partially accurate.

4.3.1.2 Spatial-temporal measures
Response profiles for the flow of liquid and DTV are shown in Figure 18. For flow of liquid, the pattern of scores was similar to that found in the previous study: the number of children obtaining perfect scores of 12 increased with age – 18 children (50%) in Y1; 26 (57.8%) in Y3; and 37 (86%) in Y5 – and there was a smaller tail of those who made one or more errors. However, the six-step version was not harder. One-way ANOVA showed age-related progression, $F(2,121)=6.625$, $p=.002$, partial eta squared=.099, with Y1 and Y3 equivalent, but lower than Y5, in line with the later development seen on the causal performance.
(a) Flow of liquid

School year

Number of children achieving

Mean = 8.3(4.3)  
Mean = 9.2(3.7)  
Mean = 11.2(2.4)

(b) Distance

School year

Number of children achieving

Mean = 4.3(1.5)  
Mean = 4.9(1.2)  
Mean = 5.2(1.1)
Figure 18. Profile of scores on (a) flow of liquid (max=12), (b) distance, (c) time, and (d) velocity (max=6 for all)
Although fewer children achieved maximum scores, performance on the DTV tasks was similar to flow of liquid. There was the same increase with age in high scores and reducing tail of those making more errors. This was most obvious for distance. Time and particularly velocity showed broader distributions of scores, suggesting that time and velocity tasks were slightly more taxing for children. One-way ANOVAs found significant increase with age on distance, $F(2, 121) = 5.423, p = .006$, partial eta squared = .082, with Y1 significantly different from Y5, and Y3 intermediate; but not on time or velocity, although both showed upward trends with age. Further, one-way ANOVA confirmed a significant effect of age on the total DTV score $F(2, 121) = 4.117, p = .019$, partial eta squared = .064, with Y1 mean = 12.58 (sd = 3.5), significantly different from Y5 mean = 14.56 (sd = 2.8); and Y3 mean = 13.33 (sd = 3.1), intermediate.

The lack of increase in children’s performances on time and velocity was surprising. Moreover they were the ones correlated weakest (Table 20). Distance-time correlations were higher than distance-velocity, though they were highly significant even age was controlled for. DTV total showed the highest correlations with flow of liquid, as well as all of its components. Although being highly significant, time and flow of liquid correlations were weaker, signaling that they measured less similar competences.

Table 20. Zero order and partial correlations between spatial-temporal measures (significant associations in bold)

<table>
<thead>
<tr>
<th></th>
<th>FOL</th>
<th>Distance</th>
<th>Time</th>
<th>Velocity</th>
<th>DTV total</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOL</td>
<td>1</td>
<td>.444***</td>
<td>.339***</td>
<td>.442***</td>
<td>.547***</td>
</tr>
<tr>
<td>Distance</td>
<td>.383***</td>
<td>1</td>
<td>.438***</td>
<td>.394***</td>
<td>.799***</td>
</tr>
<tr>
<td>Time</td>
<td>.309***</td>
<td>.417***</td>
<td>1</td>
<td>.202*</td>
<td>.747***</td>
</tr>
<tr>
<td>Velocity</td>
<td>.410***</td>
<td>.363***</td>
<td>.180*</td>
<td>1</td>
<td>.698***</td>
</tr>
<tr>
<td>DTV total</td>
<td>.503***</td>
<td>.783***</td>
<td>.742***</td>
<td>.686***</td>
<td>1</td>
</tr>
</tbody>
</table>

Zero-order correlations above diagonal, partial correlations below; N=124; *p<.05, **p<.01, ***p<.001
A further mediation analysis between the three variables, taking velocity as an outcome and time as a mediator, elaborated the effects: there was no interaction between the measures, neither mediation effect. Direct effect of time was non-significant \( b=.03, SE=.00, t=.39, p=.69 \) with a small unstandardized indirect effect of \( b=.01, CI=-.11 \) to .12, confirming that none of the measures mediated the effects of others. This was the case when the mediators replaced respectively, except for distance. Indeed, distance mediated velocity at moderate level [indirect effect of distance \( b=.15, CI=.06 \) to .29]. These patterns did not change when the same analyses were run including only Year 1 children. The only difference was CI values for time relatively increased, indicating that children considered duration slightly more in their velocity judgments.

4.3.1.3 Verbal, nonverbal, and spatial ability

Table 21 shows mean scores by age group on the spatial and general ability measures. There was significant negative skew on rotation and vocabulary, and positive on block design. One-way ANOVAs by school year found significant increases with age on each: for rotation, Welch robust statistic=15.004 (df = 2, 73.799); for vocabulary, 51.303 (2, 77.705); for block design, 24.229 (2, 79.791); \( p<.001 \) for all. For rotation, Y1 and Y3 were equivalent, but significantly lower than Y5. The later growth suggests that in this sample the Tick and Tock spatial task was more discriminating in the higher elementary age range than the task of the previous study. Performances are broadly comparable to the flow of liquid and DTV tasks then.

Table 21. Mean score (sd) by age group on rotation (max=16), vocabulary and block design

<table>
<thead>
<tr>
<th></th>
<th>Y1</th>
<th>Y3</th>
<th>Y5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation</td>
<td>12.43  (2.88)</td>
<td>12.99  (2.12)</td>
<td>14.85  (1.74)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>22.48  (5.39)</td>
<td>29.05  (5.09)</td>
<td>34.01  (4.56)</td>
</tr>
<tr>
<td>Block design</td>
<td>12.43  (5.63)</td>
<td>16.45  (7.12)</td>
<td>24.01  (8.97)</td>
</tr>
</tbody>
</table>
4.3.1.4 Statistical tasks

Using standardized mean scores on each measure, the trends were compared as in the previous study (see Figure 19). Again, the developmental trend was clear, with tasks varying in difficulty. Block design was similarly difficult for children, while flow of liquid was easier. The causal and both statistical tasks were also more difficult for children than the previous study.

![Developmental trajectories](image)

**Figure 19. Developmental trajectories**

The remaining measures showed a linear trend, with marbles exhibiting the steepest gradient. One-way ANOVAs by school year found significant increases with age on all variables, using the Welch robust statistic, p<.01 in each case, except DTV, p<.05. For all measures, there were significant increases in scores from Y1 to Y5. In this sample, children exhibited marginally lower means scores for vocabulary, block design, marbles and – even allowing for the change in task – covariation. The mean scores for covariation were 4.48 (1.96) for Y1; 5.34 (2.34) for Y3; and 6.34 (1.96) for Y5 respectively. For marbles, the mean scores were 3.76 (2.99) for Y1; 5.39 (3.05) for Y3; and 5.61 (3.32) for Y5. Although covariation and marbles showed closer paths, the standard deviations (in parenthesis) were higher, indicating that children performed differently on each statistical measure.
4.3.2 What does predict children’s reasoning about causal processes?

4.3.2.1 Correlations between variables
Flow of liquid tended an exponential rather than linear relationship with the causal measures: for total causal score, R-square for linear fit = .342, for exponential fit = .357. This suggests that it was marginally more discriminating of causal performance in the upper range of scores, reflecting the different causal task and sample characteristics. The relationships for block design were logarithmic as before.

The total causal score showed positive associations with all the predictors (Table 22). The main predictors were also all positively correlated with each other. In this sample, level of parental education – but not parental occupation – was also weakly correlated with causal performance, however it was unrelated to any of the other predictors.

Table 22. Zero-order and partial correlations between measures (significant associations in bold)

<table>
<thead>
<tr>
<th></th>
<th>Causal total</th>
<th>WASI vocabulary</th>
<th>Block design (logarithmic)</th>
<th>Flow of liquid (exponential)</th>
<th>DTV distance</th>
<th>DTV time</th>
<th>DTV velocity</th>
<th>DTV total</th>
<th>Rotation</th>
<th>Parent education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal total</td>
<td>1</td>
<td>.53***</td>
<td>.48***</td>
<td>.60***</td>
<td>.35***</td>
<td>.29***</td>
<td>.34***</td>
<td>.44***</td>
<td>.33**</td>
<td>.19*</td>
</tr>
<tr>
<td>Description</td>
<td>-</td>
<td>.52***</td>
<td>.30**</td>
<td>.36***</td>
<td>.34***</td>
<td>.26**</td>
<td>.23*</td>
<td>.37**</td>
<td>.24**</td>
<td>.01</td>
</tr>
<tr>
<td>Prediction (judgement)</td>
<td>-</td>
<td>.14</td>
<td>.25**</td>
<td>.32***</td>
<td>.17*</td>
<td>.18*</td>
<td>.12</td>
<td>.23**</td>
<td>.20*</td>
<td>.10</td>
</tr>
<tr>
<td>Explanation</td>
<td>-</td>
<td>.55***</td>
<td>.48***</td>
<td>.60***</td>
<td>.38***</td>
<td>.32***</td>
<td>.40***</td>
<td>.49***</td>
<td>.29**</td>
<td>.18*</td>
</tr>
<tr>
<td>WASI vocabulary</td>
<td>-</td>
<td>1</td>
<td>.53***</td>
<td>.48***</td>
<td>.39***</td>
<td>.25***</td>
<td>.25***</td>
<td>.38***</td>
<td>.43**</td>
<td>.01</td>
</tr>
<tr>
<td>Block design</td>
<td>-</td>
<td>.53***</td>
<td>1</td>
<td>.49***</td>
<td>.42***</td>
<td>.21**</td>
<td>.35***</td>
<td>.42***</td>
<td>.45**</td>
<td>.07</td>
</tr>
</tbody>
</table>

(logarithmic)
### Table 23

<table>
<thead>
<tr>
<th>Variable</th>
<th>.41***</th>
<th>.48***</th>
<th>.49***</th>
<th>1</th>
<th>.39***</th>
<th>.36***</th>
<th>.46***</th>
<th>.54***</th>
<th>.34*</th>
<th>.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow of liquid (exponential)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DTV distance</td>
<td>.10</td>
<td>.39***</td>
<td>.42***</td>
<td>.39***</td>
<td>1</td>
<td>.44***</td>
<td>.39***</td>
<td>.79***</td>
<td>.37**</td>
<td>.13</td>
</tr>
<tr>
<td>DTV time</td>
<td>.16</td>
<td>.24**</td>
<td>.21*</td>
<td>.36***</td>
<td>.44**</td>
<td>1</td>
<td>.20*</td>
<td>.74***</td>
<td>.35**</td>
<td>-.05</td>
</tr>
<tr>
<td>DTV velocity</td>
<td>.18*</td>
<td>.24**</td>
<td>.35***</td>
<td>.45***</td>
<td>.39**</td>
<td>.20*</td>
<td>1</td>
<td>.69***</td>
<td>.26**</td>
<td>.07</td>
</tr>
<tr>
<td>DTV total</td>
<td>.23*</td>
<td>.38***</td>
<td>.42***</td>
<td>.38***</td>
<td>.80***</td>
<td>.75***</td>
<td>.70***</td>
<td>1</td>
<td>.44**</td>
<td>.05</td>
</tr>
<tr>
<td>Rotation</td>
<td>.09</td>
<td>.43***</td>
<td>.45***</td>
<td>.34***</td>
<td>.37***</td>
<td>.35***</td>
<td>.27***</td>
<td>.44***</td>
<td>1</td>
<td>-.01</td>
</tr>
<tr>
<td>Marbles</td>
<td>.14</td>
<td>.48***</td>
<td>.50***</td>
<td>.51***</td>
<td>.38***</td>
<td>.23*</td>
<td>.39***</td>
<td>.44**</td>
<td>.25**</td>
<td>.22*</td>
</tr>
<tr>
<td>Covariation</td>
<td>.25*</td>
<td>.43***</td>
<td>.32***</td>
<td>.44***</td>
<td>.31***</td>
<td>.24**</td>
<td>.34***</td>
<td>.39***</td>
<td>.35**</td>
<td>.20*</td>
</tr>
<tr>
<td>Parental education</td>
<td>.17</td>
<td>.01</td>
<td>.07</td>
<td>.11</td>
<td>.13</td>
<td>-.05</td>
<td>.07</td>
<td>.02</td>
<td>-.01</td>
<td>1</td>
</tr>
</tbody>
</table>

Zero-order correlations, with partial correlations presented in the first column; N=124; *p<.05, **p<.01, ***p<.001

Controlling for age in months, vocabulary and log block design, the only significant predictors of causal score (as well as explanation and description) were flow of liquid, and now DTV -the other spatial-temporal measure-, and covariation. Parental education was marginal, p=.053. DTV also correlated with rotation.

### 4.3.2.2 Hierarchical regression models

The unique variance accounted for by the predictors examined with total causal score as dependent variable. Predictors were entered in the same order to the previous study (cf. chapter three), except that DTV was replaced by the speed task. The analysis produced significant R-squared change at every stage except the third, when rotation was entered (see Table 23).
Table 23. Hierarchical regression with total causal score as dependent variable (significant predictors in bold)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Model</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in months</td>
<td>-1.36</td>
<td>-2.45*</td>
<td>-2.23*</td>
<td>-2.16*</td>
<td>-2.24*</td>
<td>-2.21*</td>
<td></td>
</tr>
<tr>
<td>WASI vocabulary</td>
<td>.624***</td>
<td>.519***</td>
<td>.460***</td>
<td>.373***</td>
<td>.370***</td>
<td>.348***</td>
<td></td>
</tr>
<tr>
<td>Block design (log)</td>
<td>.336***</td>
<td>.268**</td>
<td>.185*</td>
<td>.178*</td>
<td>.193*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DTV total</td>
<td></td>
<td>.208*</td>
<td>.078</td>
<td>.068</td>
<td>.057</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOL (exp)</td>
<td></td>
<td></td>
<td>.363***</td>
<td>.363***</td>
<td>.349***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotation</td>
<td></td>
<td>.017</td>
<td>.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marbles</td>
<td></td>
<td></td>
<td></td>
<td>.050</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covariation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.128</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AdjRsquare = .490; ΔR² = .292*** for M1; .075*** for M2; .034* for M3; .077*** for M4; .001 for M5; and .012 for M6. *p < .05. **p < .01. ***p < .001.

Age was a negative predictor throughout, becoming significant from the second stage and onwards. This seems to be a function of its relationship to residual variance not explained by other predictors, since it was positively associated with causal scores (r=.285, p=.001). In contrast to the previous study, vocabulary was a strong predictor throughout, but its beta dropped at each successive stage, especially when log block design was included. Of course the inclusion of the new predictors changed the degree of freedom, but beyond that in this more diverse sample the predictors seem to share some extent of variance in predicting causal performance. Block design was a significant predictor from the second model onward. However, the inclusion of DTV and flow of liquid dropped its beta, suggesting shared variance between them.

The analysis confirmed that the new spatial-temporal measure, DTV, explained unique variance, independent of vocabulary and nonverbal ability (Model 3). However, its beta dropped sharply when flow of liquid was entered, confirming the major overlaps with it.
in explained variance. Flow of liquid was a stronger predictor of causal performance, explaining a further 7.7% of variance when it was included in the model. Note that the explained variance is here is higher than the previous study, with a larger final beta (.349). In total, the two spatial-temporal measures explained 10.5% of variance. Rotation had no predictive value, as in the previous study.

Although initially covariation and marbles were included in separate models none of them was predictive of causal total measure, as in the previous study. Including them in the same model did not cause any changes. The inclusion of the statistical measures decreased the beta of flow of liquid substantially, while this inclusion increased the beta of block design.

The same model was also used to examine explained variance in each of the component scores (description, prediction, justification, explanation), and similar results for explanation. The DTV and flow of liquid remained clear, but they shared their variances in all the models, except for the justification. For this component, both spatial-temporal measures became significant at last model. Vocabulary was generally strong, and log block design was weaker (Table 24). The models were also rerun with parental education at the final stage to check whether it added predictive power. It was a significant predictor only for justification, with a beta of .158, p=.036.

For description, the final model had adjRsquare=.329, which was acceptable. Vocabulary was again a significant predictor from the start, but the inclusion of other variables reduced its beta from to .471, p<.001 at the end. DTV was also a significant predictor when it was entered, with a beta of .205, p=.020, explaining an additional 3.2% of variance. However, it became marginal, p=.057, when exponential flow of liquid was included, though that itself was not significant in this instance. Note that description and prediction had substantially lower variance than the other causal components, which may have affected outcomes.
For prediction, the final model had adjRsquare=.158, smaller than for the other components and indicating an inadequate explained variance. In contrast to description, vocabulary was never a significant predictor at any stage. Log block design was a significant predictor until DTV was included, though that was not significant itself. Exponential flow of liquid was the only significant predictor, explaining a further 3.7% of variance, and with a final beta of .263, p=.027. Its inclusion also further reduced the beta for log block design. However, the results for the prediction component are all tentative, as the adjRsqared value of the model is not at satisfactory level.

Table 24. Hierarchical regression with components as dependent variables (significant predictors in bold)

<table>
<thead>
<tr>
<th>Model</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictor</td>
<td>β</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age in months</td>
<td>-0.095</td>
<td>-0.114</td>
<td>-0.093</td>
<td>-0.091</td>
<td>-0.086</td>
<td>-0.087</td>
</tr>
<tr>
<td>WASI vocabulary</td>
<td><strong>.588</strong>*</td>
<td><strong>.570</strong>*</td>
<td><strong>.512</strong>*</td>
<td><strong>.495</strong>*</td>
<td><strong>.497</strong>*</td>
<td><strong>.471</strong>*</td>
</tr>
<tr>
<td>Block design (log)</td>
<td>.059</td>
<td>.008</td>
<td>.025</td>
<td>.020</td>
<td>.016</td>
<td></td>
</tr>
<tr>
<td>DTV total</td>
<td></td>
<td><strong>.205</strong>*</td>
<td>.180</td>
<td>.187</td>
<td>.170</td>
<td></td>
</tr>
<tr>
<td>Flow of liquid (exp)</td>
<td></td>
<td>.071</td>
<td>.071</td>
<td>.047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotation</td>
<td></td>
<td>.026</td>
<td></td>
<td>.042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marbles</td>
<td></td>
<td></td>
<td>-0.010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covariation</td>
<td></td>
<td></td>
<td>.127</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AdjRsquare = .329; ΔR² = .279*** for M1; .002 for M2; .033* for M3; .003 for M4; .000 for M5; .011 for M6. *p &lt; .05. **p &lt; .01. ***p &lt; .001.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Prediction |       |       |       |       |       |       |
| Age in months | .004 | -0.081 | -0.066 | -0.061 | -0.078 | -0.063 |
| WASI vocabulary | .147 | .065 | .022 | .038 | .046 | .056 |
### Justification

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in months</td>
<td>-.142</td>
<td>-.215</td>
<td>-.217</td>
<td><strong>-.210</strong></td>
<td><strong>-.232</strong></td>
</tr>
<tr>
<td>WASI vocabulary</td>
<td><strong>.583</strong>*</td>
<td><strong>.512</strong>*</td>
<td><strong>.515</strong>*</td>
<td><strong>.430</strong>*</td>
<td><strong>.419</strong>*</td>
</tr>
<tr>
<td>Block design (log)</td>
<td><strong>.227</strong></td>
<td><strong>.232</strong></td>
<td>.149</td>
<td>.128</td>
<td>.097</td>
</tr>
<tr>
<td>DTV total</td>
<td>.013</td>
<td>-.141</td>
<td>-.173</td>
<td>-.199**</td>
<td></td>
</tr>
<tr>
<td>Flow of liquid (exp)</td>
<td><strong>.358</strong>*</td>
<td><strong>.360</strong>*</td>
<td><strong>.324</strong>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotation</td>
<td>.113</td>
<td>.121</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marbles</td>
<td>-.179</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covariation</td>
<td>.153</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AdjRsquare = .158; $\Delta R^2 = .022$ for M1; .046* for M2; .018 for M3; .037* for M4; .005 for M5; .030 for M6. *$p < .05$. **$p < .01$. ***$p < .001$. 

### Explanation

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in months</td>
<td>-.174</td>
<td>-.285**</td>
<td>-.258*</td>
<td>-.252*</td>
<td>-.242*</td>
</tr>
<tr>
<td>WASI vocabulary</td>
<td><strong>.667</strong>*</td>
<td><strong>.560</strong>*</td>
<td><strong>.485</strong>*</td>
<td><strong>.407</strong>*</td>
<td><strong>.411</strong>*</td>
</tr>
<tr>
<td>Block design (log)</td>
<td><strong>.344</strong>*</td>
<td><strong>.258</strong></td>
<td><strong>.182</strong></td>
<td><strong>.192</strong></td>
<td><strong>.190</strong></td>
</tr>
<tr>
<td>DTV total</td>
<td><strong>.263</strong>*</td>
<td>.146</td>
<td>.160</td>
<td>.152</td>
<td></td>
</tr>
</tbody>
</table>
Flow of liquid (exp) | .329*** | .328*** | .317***
Rotation | .049 | .061
Marbles | . | -.022
Covariation | . | .077

AdjRsquare = .529; ΔR² = .318*** for M1; .078*** for M2; .054** for M3; .063*** for M4; .002 for M5; .004 for M6. *p < .05. **p < .01. ***p < .001.

For justification, the final model had adjRsquare = .375. Vocabulary was a significant predictor throughout, but the inclusion of log block design and exponential flow of liquid reduced its initial beta, and its final value was .392, p < .001. Age became a significant predictor, again with a negative beta, following the inclusion flow of liquid, consistent with it essentially picking up residual variance. Log block design was a significant predictor until the inclusion of exponential flow of liquid. Flow of liquid was a significant predictor again, explaining an additional 7.5% of variance, with a final beta of .324, p < .001. The other spatial-temporal measure DTV became significant at the final model, but with a negative beta, indicating that some children’s DTV performances in young age group significantly affected the outcomes beyond the age effect. Neither rotation nor statistical measures were significant.

For explanation, the final model had adjRsquare = .529, with the highest explained variance. This model was almost identical to the causal total outcomes. Vocabulary was a significant predictor throughout, but its beta dropped with the inclusion of log block design and the spatial-temporal measures to a final value of .398, p < .001. Log block design was significant once it was entered, though its beta also reduced when the spatial-temporal measures were included, with a final value of .190, p = .028. Age was a negative predictor, as for overall causal measure, final beta = -.241, p = .012. DTV was a more substantial predictor of explanation than the other components, explaining 5.4% of variance when it was included, with a beta of .263, p = .001. However, it became marginal, beta = .152, p = .054, when exponential flow of liquid was included. Flow of liquid was once more a substantial predictor, explaining 6.3% of variance with a final beta of .317,
p<.001. None of the spatial or statistical measures survived in the model, but their inclusion slightly reduced the betas of the other variables.

4.3.2.3 Path analysis
The regression analysis suggested that flow of liquid and vocabulary, and to a lesser extent block design, were direct predictors of causal performance in this sample. But also, from the pattern of reduction in beta values, block design was a prior influence on (i.e. partially mediated by) DTV. Both had a similar relationship to flow of liquid. Age picked up residual variance not accounted for by these. This pattern of effects is captured by the extended model shown in Figure 20. Path analysis uses the maximum likelihood approach, which also confirmed the extended model providing a high degree of fit to the data, chi-square = 0.953, df = 1, p = .329.

Figure 20. Path model including standardised coefficients for the effects of exponential flow of liquid, DTV, log block design, age and vocabulary on causal total score

This model differs from the previous model (cf. chapter three) and reflects the methodological changes between the two studies:

(1) The more diverse sample of the replication study led to a stronger relationship between vocabulary and causal performance.
(2) In the recent model, the effect of log block design was partially mediated by flow of liquid, detectible due to the inclusion of DTV.

(3) There was also now a strong link between DTV and flow of liquid.

(4) Replicated in both models, there are strong direct effects of spatial-temporal and nonverbal ability on causal performance.

4.3.2.4 Nature of shared variances between predictors

Factor analysis with varimax rotation (KMO=.864, p<.001) clarified the nature of the shared variances between vocabulary, flow of liquid, DTV, log block design, marbles, and covariation. A four-factor model provided the clearest solution, with the first factor explaining 24% of variance, the second 23%, the third 21%, and the fourth 18% (Table 25). This confirmed log block design as being most closely related to marbles, flow of liquid to DTV, but covariation as being distinct, though block design shared a substantial variance with vocabulary too.

Table 25. Three factor solution for the relationship between exponential flow of liquid, DTV, log block design, marbles, and covariation

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp FOL</td>
<td>.646</td>
<td>.375</td>
<td>.292</td>
<td>.221</td>
</tr>
<tr>
<td>DTV total</td>
<td>.912</td>
<td>.151</td>
<td>.139</td>
<td>.148</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.197</td>
<td>.167</td>
<td>.884</td>
<td>.259</td>
</tr>
<tr>
<td>LNBlock</td>
<td>.254</td>
<td>.603</td>
<td>.583</td>
<td>-.015</td>
</tr>
<tr>
<td>Marbles</td>
<td>.229</td>
<td>.881</td>
<td>.154</td>
<td>.231</td>
</tr>
<tr>
<td>Covariation</td>
<td>.214</td>
<td>.171</td>
<td>.192</td>
<td>.926</td>
</tr>
</tbody>
</table>

Taking exponential flow of liquid as standing for both spatial-temporal measures, further maximum likelihood path analysis was used to examine whether either the extended mediation model, or a reversed version of this (i.e. with covariation and marbles swapping position) provided an adequate fit to the data. These models were contrasted
with a further extension of this, in which log block design and marbles fed directly into flow of liquid alongside covariation, reflecting the somewhat more balanced influence of the two statistical measures.

Figure 21. Further extended mediation model for the effects of log block design, covariation, marbles and flow of liquid on overall causal performance.

For both causal indices –inference and causal total–, the further extended model provided the best fit to the data: in each case $\chi^2 = 0.315$, $p=.575$, df=1. Figure 21 illustrates the model plus path coefficients obtained for overall causal performance. Again then, there appears to be a stable pattern of effects in which nonverbal ability, covariation and probability all support spatial-temporal analysis, but with each also influencing specific aspects of causal thinking directly. In this sample, flow of liquid and vocabulary have the strongest effects for both causal measures: for inference, the effects of flow of liquid and vocabulary are .36 and .41, and those of nonverbal ability, probability and covariation ranged as .20, .01, and .09 respectively.

Further moderation analyses indicated no interaction effect between log block design, marbles and flow of liquid in predicting causal performance [For log block design R-squared=.41, $p=.42$; for marbles R-squared=.38, $p=.15$; and here there was no conditional
effect of covariation, $R^2=.41$, $p=.15$. For log block design and marbles, explained variance was relatively low, and there was no interaction between them, $R^2=.26$, $p=.65$.

4.3.3 Identification of causal mechanisms

As in the previous study, the specific relationship of flow of liquid to inference of mechanisms was examined. Table 26 shows the percentage of children obtaining perfect scores for flow of liquid at each level of explanation score for sinking, absorption and solution. The pattern here was clearer than the previous study, probably assisted by the refinement of the scoring system. For all three processes, performance on flow of liquid was strongly associated with coordination of variables, and the step from there to mechanism was always made by those who 'got' it.

Table 26. Percentages of children obtaining perfect scores on the flow of liquid task at each level of explanation response

<table>
<thead>
<tr>
<th>Causal task</th>
<th>Inference score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4 (mechanism)</td>
</tr>
<tr>
<td>Sinking</td>
<td>100.0</td>
</tr>
<tr>
<td>Absorption</td>
<td>100.0</td>
</tr>
<tr>
<td>Solution</td>
<td>100.0</td>
</tr>
</tbody>
</table>

While perfect performance on flow of liquid did not guarantee coordination of variables or inference of mechanism, the percentages of children who demonstrated such performance declined sharply below the level of coordination responses. This suggests that if they failed to grasp flow of liquid, they were unlikely to achieve coordination, neither inference of mechanism. Supporting these, the chi-square results for explanation score by perfect versus non-perfect flow of liquid score were significant for sinking, chi-square=31.496; absorption, chi-square=21.882; and solution, chi-square=38.717; df=4,
p<.001 for all. All of those who gave any mechanism responses had a perfect flow of liquid score; of those who gave none, it was 59.0%, Fisher's exact test, p<.001.

A similar pattern appeared for DTV. Children with coordination and mechanism responses consistently had significantly higher scores on this measure too (Table 27): comparing those who made any mechanism responses versus to those who did not, mean = 15.89 vs. 13.11, t(1,122)=5.025, p<.001. Further one-way-ANOVAs for sinking, F=7.660, partial eta-squared=.205; for absorption, F=7.771, partial eta-squared=.207; for solution, F=7.756, partial eta-squared=.207; df=4,119, p<.001 for all. All confirming that causal inference and awareness of invisible mechanisms is strongly related to the forms of spatial-temporal analysis as indexed by both flow of liquid and DTV.

Table 27. Mean DTV scores (sd) at each level of inference response

<table>
<thead>
<tr>
<th>Causal task</th>
<th>Inference score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4 (mechanism)</td>
</tr>
<tr>
<td>Sinking</td>
<td>15.40 (3.13)</td>
</tr>
<tr>
<td>Absorption</td>
<td>15.20 (1.30)</td>
</tr>
<tr>
<td>Solution</td>
<td>16.27 (1.55)</td>
</tr>
</tbody>
</table>

4.3.4 *Significance of correlations in two unequal samples in chapter three and four*

The correlations between spatial-temporal and causal measures differed depending on the characteristics of the samples in two studies [e.g. Flow of liquid rstudy1=.52 (N 107); and rstudy2=.60 (N 124)]. Using Fisher’s transformation, z values showed that the difference between the two correlations was not significant (z=±.875 for causal total with a probability of .192). Because the probability of the observed z score is greater than .05 (.192>.05) it implies the correlations between FOL and causal total in both studies are similar, and the observed variability is due to chance. Similarly, neither vocabulary (z=.073) nor block design (z=1.34) z values were significant.
Comparatively, differences in correlations between marbles and causal total in two unequal samples were explored further \[ r_{\text{study1}} = .55 (N = 107); \text{and} \ r_{\text{study2}} = .39 (N = 124) \]. Using Fisher’s transformation again, the z value indicated that the difference between two correlations was not significant \[ z = 1.54 \text{ for causal total, with a probability of } .06 \]. Since the level of significance is set at .05, and the probability of the observed z score is greater (.06 > .05), this also implies the correlations between marbles and causal measure in both studies are similar.

4.4 Discussion

This chapter concerned whether the findings in Chapter three were replicable, aiming to explore further the link between children’s reasoning about causal processes and the use of probability, covariation, and spatial-temporal information. The modified causal tasks followed the structure of a scientific investigation; an extended flow of liquid task; an additional spatial-temporal measure distance/time/velocity integration tasks; and a more socially representative sample within the same age range were employed.

Two measures assessed statistical ability, one addressing probability (the marbles task), and another covariation task. These were selected based on their relative predictive strength in the previous study. The covariation task was a revision of that used before (cf. chapter three); it also utilized physical materials in the form of decks of cards rather than a computer display, in keeping with most of the other test materials. The cards task was dropped, in view of its overlap with marbles in Study 1, and randomness was dropped because of its lack of predictive power.

This study confirmed and extended the key findings of the previous study. First, as before, children gave accurate descriptions of what they observed in the causal tasks, but their ability to predict, justify and explain what they witnessed – especially with regard to mechanisms – increased more slowly, although a reasonable percentage were able to give explanation responses that explicitly coordinated variables. Except for permitting this additional level of inference, increasing the number of observations did not increase the pattern of performance noted in previous study.
Despite differences in the sample, this study confirmed the developmental trends observed previously, with a clearer picture showing the increases across the age groups. The ability to analyse spatial-temporal transformations also improved over the age range, with children exhibiting similar performance on both flow of liquid and the new DTV tasks.

This time, the DTV was a good predictor of explanation and overall causal performance, explaining unique variance in both. It shared its variance with flow of liquid – the stronger predictor – when that was included. This suggests that the common elements of both tasks were the sources of the influence on causal performance: ability to reconstruct a sequence and to mentally simulate distance-time-velocity consistently predicted higher levels of causal inference – both coordination of variables and reporting of mechanisms.

Despite differences in (1) methodologies, (2) sample characteristics, and (3) similarly discriminating, the spatial measure rotation had no predictive value; again it shared its variance with the verbal measure very early on. In this more diverse sample however, vocabulary, i.e., generic verbal ability, was a strong predictor of performance overall. Nonverbal ability had a weaker direct influence, but had an additional effect via its impact on spatial-temporal ability.

Neither statistical measure was a significant predictor in the models. Their effects were taken by the blocks and spatial-temporal measures. Covariation was marginally the stronger predictor here, and more related to blocks – possibly reflecting the revised measure used here. The reduced influence of probability and covariation may be a reflection of the less developed nature of both competences – and causal inference – in this relatively low achieving sample.

Regarding the hypotheses, it can be stated that this study largely replicated the results of the previous study, while extending the effects, showing that the network of interrelated influences on both spatial-temporal analysis and reasoning about causal processes is
based on partially unique contributions from nonverbal ability. Although various differences, both spatial-temporal measures were stronger predictors of reasoning about causal processes, indicating their unique variances were not dependent on any of the measures used in both studies – consistent with the factor analysis results.

Given the characteristics of the causal tasks, statistical ability may have constrains in directly predicting single causal phenomenon. In line with the results from the previous study, the influence of probability on inference was weaker, covariation was a more stronger variable, and they weakly associated to each other, suggesting that in this lower performing sample understanding of probability did not play a distinctive role in thinking about causal processes.

4.5 General discussion

4.5.1 Developmental trajectories

Children’s responses showed clear progress with age on all tasks. Both studies provided consistent evidence that primary schoolers have the ability to think about continuous causal processes, and also demonstrated that constrains on their performance are not a function of their immediate prior exposure to instances of these. The different task components varied in difficulty, however: around 80% of children across all ages, tasks, and in both studies made and reported completely accurate observations. For prediction, the percentage decreased to just under 50% overall. Explanation scores trailed, with only about 15% reaching top scores (lower than the previous score of 19.6%).

Predictions in previous study were made ahead of observation, so relied highly on prior knowledge, either consolidated from past observations or others’ reports, plus recognition of how this applies to the present context. This was found to be more demanding than description of current observation. The sizable difference in performance suggests many children lacked well-developed existing knowledge for any of the three processes. In replication study, prediction scores were made after initial observations to inform prediction, but performance remained poor. This could reflect sample differences, and also the increased complexity of the task.
As compared to prediction and description, children’s explanations were the lowest at all ages in both studies, and mechanism responses were relatively unusual. In the previous study, only 20.6% of children, the majority in the oldest age group, identified an underlying causal mechanism, with 5.6% doing so across more than one phenomenon; the lower corresponding values in replication study were 15.3% and 1.6%, again consistent with the more socially mixed sample. Across both studies, children’s explanations were typically limited to abstracting causal variables from the observed objects. At the same time, the more extended range of examples employed in replication revealed that a reasonable proportion of older children had progressed not only to linking causal factors to observed effects, but also to recognizing the coordinated operation of key variables, i.e., to recognizing that there is an underlying relationship. This in turn may lead into thinking of how the variables are connected to each other and to the effect, as an important precursor to thinking about mechanisms.

Mechanism level inferences remained less common in contrast to past research showing that pre-schoolers are able to make inferences about mechanisms in simpler machine/toy systems (cf. Bullock et al., 1982; Buchanan & Sobel, 2011; Shultz, 1982; Schlottmann, 1999). The argument is that this striking lag is because continuous processes might be hard for children, since they require spatial-temporal information to be analysed in different causal systems, as discussed before.

In previous study, there was greater growth on the causal and spatial-temporal tasks between Y1 and Y3 than between Y3 and Y5. In replication, this pattern was the same for the spatial-temporal measures, but reversed for the causal tasks, with the growth observed between Y3 and Y5. Given that the methodology was different in both studies, direct comparisons are not possible. However, looking at children’s explanations, it can be said that explanations restricted at all ages, and even in Y5 they seemed to find it difficult to progress to explaining the mechanisms mediating cause-effect relationships, focusing on more observable and salient factors and variables. However, this pattern was worse in the replication study as discussed earlier.
The clear impact of verbal ability in this study seems to be attributable to the more mixed sample. This may reflect vocabulary having a greater influence at lower levels of ability. The high correlations between vocabulary and block design across both studies were persistent. However, in previous study children’s performance on vocabulary was above average level at international contexts, distributing the sample across a narrower band. In the replication, the band was wider, as children’s verbal scores varied widely. There was a tandem between verbal and nonverbal ability in a way that when verbal ability spanned across a narrower band, nonverbal ability became more discriminating. When the verbal ability varied widely, it became more discriminating, though along with the nonverbal ability.

In contrast, performance on the liquid flow task in particular approached ceiling by Y5 in both the five- and six-step version, regardless of sample differences. However, the other spatial-temporal task, speed, was more difficult for even older children. The nature of the task might play role on this. As discussed in chapter three and also see Appendix 1, this task computerized spatial-temporal qualities in a highly perceptual protocol unlike analytical flow of liquid. Children’s performances varied highly, did not offer any consistent pattern across the sample (e.g. high achievers, or low achievers). Speed results implied that the ability to deal with the type of spatial-temporal manipulations required other cognitive resources beyond attention. This point investigated by the DTV here, which required mental representations to be utilized in various forms ranging from perceptual to imagery. Performance on all DTV measures and flow of liquid showed similar increases across the age range, with most children performing well on all these tasks by Y5, indicating they were comparable in terms of difficulty. Correlations on flow of liquid and DTV were higher, in contrast to flow of liquid and speed in the previous study, even when age and general ability were controlled. This may be the consequence of DTV, in particular distance, time, and flow of liquid tapping into a common underlying dimension of analytic – as opposed to pure perceptual – spatial-temporal ability, while velocity is employing more imagery. Shared variances in the regressions were also confirmatory.
One can suggest that while flow of liquid clearly involves a sequence of spatial-temporal state changes, it may draw on children’s logical ability to form a transitive, ordered series, visual-spatial working memory capacities, or knowledge of a causal process (e.g. flow under gravity, even if the response did not require any articulation of this causality). This clearly merits a further investigation, but for now what this study tells us that none of the DTV measures relies on transitive inference or causal analyses, and the high correlations between the flow of liquid and DTV tasks suggest substantial overlaps. These high correlations highlight that the kinds of spatial-temporal elements in both tasks are similar and substantially central.

We also need to acknowledge that working memory capacity may influence performance on both tasks. Given that these tasks rely substantively on manipulations of visual-spatial and temporal elements, memory effect might be inevitable in task performance. However, a preliminary research conducted by Lim (2019) suggests that any such influence is a modest and peripheral one. In his thesis, Lim used the Corsi Block Span and found that working memory explained a little over 10% of variance in performance on the flow of liquid task in a sample of 10 to 13 year olds, and was not predictive of causal inference.

Growth in statistical thinking appeared to be slower than in spatial-temporal ability, but faster than in causal inference. Past research examining children’s and adult’s covariation and probabilistic thinking in causation – Bayesian and causal learning literature – has focused on identification of the structure of causal relations between distinct variables (i.e. the relationship between use of aspirin and headache), or the strength between causes and effects based on summaries of repeated observations (i.e. the degree to which aspirin alleviates headaches), or compared children’s understanding of common cause and causal chain structures (see e.g. Lagnado et al., 2007; McCormack et al., 2015, 2016). Although no data available on direct comparability, the statistical tasks here showed similar differences in children’s statistical thinking that are largely in line with the results from those.

On the covariation task, children appeared to progress with age in their ability to assess co-occurrences, the results refining by year five. Regarding the numerical quantification,
this competence appeared to develop late, even in year five children. This is in some ways consistent with Shaklee and Mims’s (1981) finding. Although using a causal event approach, he found that children’s strategies for addressing covariation increased in complexity with age. Correspondingly, in this study older children performed better on the computation of estimates of associational strength. The data here also suggest that progress may be slower where the concern is frequency relationships, rather than frequency relations (i.e. not two cells of the 2x2 contingency table) between cause and effect, probably because the former is more abstract.

On both versions of the covariation tasks, one item provided participants with a hundred percent co-occurrence information, which made the interpretation of stimuli unambiguous. The following presented incomplete information, which increased the ambiguity from 75%, in replication to 50%, and to 0%. Children mostly seemed to deal well with hundred percent co-occurrence (perfect covariation fail rate 9% in first study; it is 14% in replication), but less well with the zero correlation (fail rate 34% in first study; it is 39% in replication), especially in the youngest age group. This is consistent with Koerber et al.’s (2005) finding that even older children show difficulties in interpreting instances of non-covariation between two distinct events. However, unlike that study, the tasks here did not include a conflict between previous beliefs and causal evidence requiring children to test their hypotheses, only to interpret imperfect covariation patterns in a non-causal context. The consistent difficulty in interpreting non-covariation data in both approaches suggests a more fundamental problem that requires further investigation.

In the probability tasks, young children did not show a clear numerical grasp of probability. Majority of Year 1 and Year 3 children preferred to tick the line on the score sheet rather than giving a numerical answer even if they were asked. They did show some understanding of possibilities and their thinking on these tasks seemed to be binary: they focused on whether there was a good chance of winning or losing, rather than the degree of that chance. The results therefore resonate with Piaget and Inhelder’s (1951/1975), as well as White’s (2014) findings (see also Fernbach & Sloman, 2009; Meder, Gerstenberg, Hagmayer, & Waldmann, 2010; Rottman & Keil, 2012; and McCormack et al.’s 2015, 2016), adding to those that thinking with numbers and computing probabilities appeared
to start from Y3 onwards, consistent with the covariation findings. In neither tasks children began to approach ceiling performance, even by Y5.

There were also departures from past findings. In particular, many of the younger children even in the higher performing previous study sample did not find it easy to make the distinction between predictable and random events. The randomness task there showed substantial variation around the mean of 1, but with a high standard deviation indicating that majority of children were scoring around zero in Year 1, and growth coming predominantly between Y3 and Y5. This seems to be contrary to Kuzmak & Gelman’s (1984) study at first glance, but the results need to be interpreted carefully for two reasons: first, the variance in scoring of the randomness task was quite limited, second it has no predictive power. The outcome can be task specific then. Conversely, children did not find it harder to deal with sequential as opposed to classical probability, as also suggested by work on probability learning (e.g. Brainerd, 1981). The cards task showed very similar developmental profile to the marbles task and they were strongly correlated. On both tasks, there was a general improvement of probabilistic thinking through the primary age range. Children seemed to move from awareness of/sensitivity to variation in likelihood to numerical calculations of this. Together with the outcome of randomness, this requires a further investigation then. While recent cognitive-developmental work has focused on tasks sensitivity to not to miss out youngest children’s level of skills (Schlottmann & Wilkening, 2011) the tasks used here did not stretch young children, they were given the options to express either their feelings or proportional calculations – ratios and decimals – as opposed to more basic judgments. The scoring system was friendly to the both types of judgments. However, in both studies statistical tasks highlighted late development.

4.5.2 What predicts children’s reasoning about continuous causal processes

4.5.2.1 Spatial-temporal measures
In previous study, one of the spatial-temporal, flow of liquid task, distinguished between lower and higher levels of overall causal performance across the age range, and was the only predictor of higher levels of causal explanation. It was the only measure consistently
sharing little variance with the other predictors, including the spatial measures, confirming its distinctness from these. This result was replicated here in chapter four. Even with the extended causal task indices and measurement style, the flow of liquid spatial-temporal measure remained the strongest predictor of children’s causal performance – stronger in replication overall than in previous study, and if anything discriminating to a greater extent at the higher levels of overall causal performance.

While flow of liquid clearly involves the sequence of spatial-temporal state changes, one can suggest it may rely on children’s logical ability to form a transitive, ordered series, or on knowledge of a causal process, flow under gravity, even if the response did not require any articulation of this causality (J. Hurry with personal conversation). To clarify this point, another subsequent study with 7 to 10 year olds (N=76) compared performance on the five stage flow of liquid task with that on (1) a causal ordering task (Piaget’s causal seriation task), and (2) a non-causal memory task (an identical method to flow of liquid, requiring reordering of the sequence of presentation of five pairs of simple abstract shapes without any causal/temporal element). Controlling for age and vocabulary, flow of liquid was uncorrelated with seriation (causal ordering) (r=0.031, p=0.916), but was significantly correlated with the non-causal memory task (r=0.354, p=0.002), providing further evidence that flow of liquid measures the ability to recreate observed sequences based on spatial-temporal analysis, and this can be done without including of causal relationships. This is consistent with the high correlations between the flow of liquid and DTV tasks, which were all spatial-temporal.

Unlike the speed task, the other spatial-temporal measure, DTV, designed to share analytic characteristics with flow of liquid, was also a significant predictor of overall causal performance, coordination of variables and inference of mechanism. It too explained unique variance beyond verbal, nonverbal and spatial measures, and it correlated well with flow of liquid, controlling for other factors. This clear overlap in explained variance again indicates that the DTV tasks and flow of liquid measure similar spatial-temporal competences. However, DTV did not survive the inclusion of flow of liquid in the regression models and the path analysis indicated that the influence of DTV was subsumed by its commonality with flow of liquid, while flow of liquid appears to
assess a further and additionally predictive dimension. One obvious candidate for this would seem to be the ability to use spatial-temporal information to extract the underlying principle governing the pattern of change.

Correlations between distance, time, velocity was moderate/medium in size; with no mediation effect between them, assuring that they measured relatively dissimilar competences, except for distance, which mediated velocity at low and medium level. This contradicts with Piaget’s view, suggesting time-speed confusion is a consequence of the space-speed disarray (comparing the duration of two moving bodies, children attribute longer duration to the longer distance). In this sample children’s judgments of velocity were partially affected by distance, rather than duration (time). Velocity-time associations were consistently weak, highlighting that children attributed longer distance mostly to the speed. However, this does not falsify Piaget’s view completely, as the distance-traveled cue is still used by some other children.

In the distance and velocity tasks, most young children frequently traced the motion of the animals one by one across the screen with their finger, apparently using this as an external accompaniment to support their judgments, indicating that children needed to show great amount of attention to computerized duration intervals. Probably this was a useful method for them to determine the distance/motion as a whole, probably the position of the finger/hand aided them to analyze of the coarsest in relation to the hand. This resonates with the above mentioned idea that it’s not easy for children to approximate virtual manipulations.

It is also plausible to think that the different processing characteristics of the tasks were responsible for their differences in performance and predictive value, as discussed earlier. Reminding the primary purpose of this research, the aim here is to ascertain if reasoning about causal processes was predicted by spatial-temporal skills, the replication study therefore focused on analytical spatial-temporal ability, rather than perceptual. Thus speed task was replaced with a set of tasks (DTV), but with different processing characteristics, requiring children to make mental simulations, with an accompanying explicit time count for segmented the imagined motion states. The flow of liquid involves
an analytic-inferential response, while the speed task afforded rapid perceptual judgments. Flow of liquid may have been relatively easy for children, because of the care taken to get the initial observation correct. Before the testing it was clarified how to draw representations related to the observations. Although similar amount of care taken for the speed, it seems to be difficult for children to orient to the rapid motion. The time task in DTV was slightly similar to the speed, but required the consideration of the movement of only one object, rather than comparing the three, as in speed. This also confirms why children struggled with the speed, but not with time.

Findings here should logically support Block, Zakay, & Hancock’s (1999) attentional gate model, also Rattat’s (2010) arguments, suggesting that attentional processes to the passage of time are necessary for moderating time perception (cognitive approach). However, analytical liquid flow and imagery velocity tasks necessitated further cognitive resources beyond attention, and also beyond the recognition of conceptual changes (see e.g. Weist, 1989 for conceptual change approach underlying the role of acquisition of tense and temporal language in understanding the difference between past, present, and future situations). To elaborate, in the velocity task, children commonly discounted the slowest/fastest animals (turtle, snail vs. dear, horse) as candidates at the outset –the rate of motion changes towards a given direction-. Both points are consistent with them employing an analytical spatial-temporal approach. However, there were task-specific variations in the requirements of these two tasks. The velocity task required children to imagine and project object movement, not needed like in flow of liquid where children simply re-ordered their drawings of a directly witnessed and explicitly segmented sequence. Moreover, although scores on the three DTV tasks were broadly similar, velocity was slightly more difficult, sharing less variance with the others, but predicting causal tasks better.

Overall, even with the extended causal task and measurement indices, the flow of liquid spatial-temporal measure remained the strongest predictor of children’s causal performance –, and discriminating to a greater extent at the higher levels of overall causal performance. All together there is a robust finding here underlying the close/unique link
between the mental forms employed in spatial-temporal analysis and causal thinking. From the developmental point of view, three main conclusions can be drawn then:

(1) Spatial-temporal analysis seems to have multiple forms, which employs various cognitive resources, such as perceptual, analytic, imagery.

(2) These forms can be used for causal and non-causal circumstances, highly significant for comprehending causal processes.

(3) Children’s temporal representations seem to rely on spatial elements even at imagery level. This conclusion is parallel with Casasanto and Broditsky’s (2008); Piaget’s (1969/2006); Piaget and Inhelder’s (1971) arguments; and inconsistent with Droit-Volet, Clement, and Fayol’s (2003) views indicating that time is an independent dimension in children’s duration judgments.

4.5.2.2 Spatial measures
Of three spatial tasks across the two studies, only one, monkey rotation, predicted one of the components, description performance, independently. The other spatial measures (paper folding, Tick and Tock) were not related to any of the causal indices or to explanation or mechanism level responses. These suggest that although spatial ability takes various forms, the impact of some of its forms (i.e. intrinsic static/dynamic and extrinsic static as indexed with the spatial tasks used here) in causal reasoning about continuous processes seem limited, and appears to overlap in particular with verbal ability. This is also contrary to Paivio’s (2013) argument discussed earlier, as in both studies spatial measures shared their variances with the verbal measure consistently. This point also merits further investigation. It is too hard to interpret why in both samples spatial measures did not share their variances with nonverbal or spatial-temporal measures, but with the verbal measure.

This lack of predictive power might be due to spatial tasks relying on two-dimensional representations and not carrying temporal information as discussed in the previous chapter. Missing the temporal element may cause them to be encoded in a discrete fashion instead of on-line spatial representations. That the spatial measures consistently shared their variances with verbal ability may suggest that mapping spatial properties into language is inherent in semantic representations as discussed by Hermer-Vazquez,
Spelke, and Katnelson (1999). This study focused on the role of natural language in spatial memory when using geometric and non-geometric information to relocate the body. And also Ratliff and Newcombe (2008) discussed whether language is necessary for spatial reorientation (see Newcombe & Shipley, 2015, for spatial thinking and its topology).

4.5.2.3 Verbal and nonverbal measures
Nonverbal ability, in contrast, as indexed by block design, did predict causal performance independently of spatial-temporal ability. However, its influence was consistently restricted to discriminating at lower levels of causal performance in both samples (i.e. the trend was consistently logarithmic). This may be because it involves the detection and analysis of spatial – but not temporal – patterns (cf. the identification of factors and variables), analytical and analogical thinking, which is then superseded by spatial-temporal analysis as driver of later development of causal thinking.

Verbal ability, as indexed by WASI vocabulary, did not predict causal performance in previous study, which had a quite homogenous sample with relatively high socioeconomic background (nonverbal ability was more discriminating in this case), but it became prominent in the replication, when the gap between high and low income families and school catchment areas widened. However, vocabulary was not associated with the parental indices in the replication, and the effect was not attributable to the sample exhibiting a greater spread of verbal ability: comparison of the means and standard deviation across the two samples shows that in previous study verbal scores clustered in the middle of the scale, and were similar to those of children in replication, possibly due to schools equalizing verbal ability. However, in replication the baseline was six points lower (min=8; max=43) than the previous study (min=14, max=43), and replication study verbal measures ranged widely. The key difference between the two samples was that in replication, children’s nonverbal ability was notably worse – unexpectedly so for children of higher educated parents, but perhaps again reflecting schooling. This led to verbal ability having greater importance for causal performance. This suggests that it is the nonverbal ability that plays the major role as a foundation for spatial-temporal ability, in line with the replication path model. It also suggests that there is some compensatory relationship between nonverbal and verbal ability, explaining why
vocabulary was as strong a predictor of the causal indices as flow of liquid. Rather than a wider effect of social area and home environment then, it may be the relationship between verbal and nonverbal ability that matters. This plainly merits further investigation.

4.5.2.4 Statistical measures

Despite the covariation and probability tasks drawing on related types of frequency information, the distinction between them was confirmed by both correlational and factor analysis in both studies. Frequentist (cards) and classical probability (marbles) tasks showed closely related performance compared to covariation and randomness tasks, and the covariation task appeared to demand a more distinct competence. In replication, marbles and covariation again remained independent, but with marbles being associated with block design, all suggesting that the ability to utilize covariation, probability and spatial-temporal information require different competences.

The predictive power of these tasks also varied. Randomness had no predictive value for either of our causal measures. As discussed earlier, this may be a function of the task and its scoring system, which varied between 0 and 2. It cannot be concluded whether basic understanding of randomness/unpredictability assists causal inference in continuous processes. This point needs further investigation with a more discriminating child-friendly task.

In previous study, marbles, in terms of overall causal performance, was a stronger predictor than covariation, but in the replication this position reversed: covariation was a stronger predictor. However, neither of them remained significant in the final models, where nonverbal and spatial-temporal measures were more dominant. In the previous study, when spatial-temporal measures were excluded, marbles predicted one of the components, explanation, indicating that it was a stronger predictor than covariation. In the replication, the same models were tested, and covariation became marginal. However, in replication none of the statistical measures survived in the final model. The reason marbles lost its predictive power for both causal measures in replication seems to be that the covariation task was more powerful and discriminating, and its effect was subsumed
by block design. Differences between the samples indexed by socioeconomic status seemed to play a role in relative developmental levels across the measures.

On the discussion whether covariation/probability information is needed for causal inference (Ahn et al., 1995; Glymour & Cheng, 1998; see also Cheng, 1997 and also Buehner & Cheng, 2005 for power PC theory which formalizes how humans infer unobservable causal powers from covariations) this study implies that the predictive power of probability and covariation tasks cannot be underestimated. Although studies did not directly measured, covariation particularly probability information must sometimes be useful for identifying candidate factors. However, they may not be as important when other kinds of information and analysis are available, such as spatial-temporal. The causal tasks here of course focused on processes rather than events or regularities. But the resonance between the results-process approach here and event approach suggesting a competition between covariation and temporal elements- (see Bullock et al., 1982; Lagnado et al., 2007; Lagnado and Sloman, 2004; McCormack et al., 2015; White, 2002) are reassuring. The literature has indicated limitations to infer causal relations from covariance alone, and that temporal cues are treated as a more stable indicator, in particular when one is intervening on a system, by introducing or removing a potential cause while keeping other factors constant to reveal the causal structure (see also Cheng, 1993; Cheng et al., 1996; Waldmann & Holyoak, 1992; Waldmann et al., 1995; and with children see Mendelson & Schulz, 1976; Siegler & Liebert, 1974; Bullock et al., 1982). Computing of covariation has also been found to be more successful when prior intuitions about the structure and temporal precedence of the causal model are involved (see Waldmann, 1996; Waldmann & Hagmayer, 2001). Although the focus is on continuity and single instances here rather than contiguity and regular data, the results suggest that statistical thinking may be promising for supporting reasoning about causal processes.

While the statistical variables were largely not significant predictors in the regressions, and path analyses for causal total measure, they still formed a network of interrelated competences along with nonverbal, verbal and spatial-temporal ability. It is the nature of this network that needs to be explained. Given the nature of the causal tasks employed,
one possibility is that the role of statistical and nonverbal ability may be enabling forms of pattern detection. In this view, block design may assess the ability to analyse and reconstruct perceptual patterns, which facilitates detection and representation of causal effects. Covariation may assess the ability to track connections, which facilitates identification of relationships between variables and outcomes. Probability may assess the ability to track the ‘definiteness’ of outcomes, which facilitates awareness of strength of effect (see also Gopnik et al., 2001; Gopnik et al., 2004; Kushnir & Gopnik, 2005, 2007; Wu et al., 2011, on the use of covariation to capture causal structures).

Framed in this way, these competences would correspond to the detection, representation and organization capacities, and then assisting or overlapping with spatial-temporal ability in analysing continuous causal processes: bootstrapping causal thinking. Each can of course have an indirect effect via spatial-temporal ability, as the correlations illustrated. However, understanding of probability seems to do something more than this proposed indirect influence on causal thinking. The correlations between explanation and the marbles in the previous study were higher, and it had the highest beta in the regression for explanation, beyond all the predictors. These suggest some other additional insight. A cross check between probability and causal task performance shows that in both studies children who had perfect marbles scores were more likely to provide high level explanation scores, and make reference to mechanisms (n=21 in previous study; n=19 in replication). Probabilistic thinking seems to be important not only for identification of the strength of the effects of variables, but may for considering unseen elements of causal processes, as suggested in the introduction. It is plausible suggest that awareness of probability drives the sensitivity to unseen factors. This would be consistent with children showing similar limitations in probability scores and references to mechanism.

Overall, considering the nature of the paradigm used in casual tasks here, nonverbal ability, probability and covariation seem to help by enabling children to identify variables and to sense that there is more to be explained about how these operate. But it seems that spatial-temporal is the primary that takes them beyond this to coordination of information and mechanism. Verbal ability also seems to be necessary, given its high influence in the
lower-performing sample, replication. Taken together, all these competences seem to have distinct and converge developmental trajectories in supporting of causal inference, but a few ones have in particular strong influence. The nature of the growth of this convergence requires further investigation. Unfortunately, we do not have a methodology to investigate this further more directly; the assumptions are restricted to correlations here.

4.5.3 The cognitive processes involved in children’s spatial-temporal analysis

Consideration the characteristics of the spatial-temporal tasks and the findings related to them provides further insight into the nature and development of spatial-temporal analysis. For instance, flow of liquid focused on the segmentation and organisation of time-dependent spatial changes, and the capacity to capture these enabled children to move beyond rapid perception of continuous causal processes. The DTV tasks provided/asked for an explicit time count that required children to mentally simulate and project the movement of a set of objects, coordinating and integrating information about other given dimensions. This requirement clearly corresponds to the demands of thinking about differences in speed of continuous causal processes. In both cases then, segmentation per se is not the only issue: upstream spatial-temporal skills also matter. These include updating/coordination of spatial and temporal information, projection of movement and extraction of underlying principles governing time-related changes.

It may not be limited to those however. In applying spatial-temporal ability to causal processes, the multiple forms of spatial-temporal thinking seem to ultimately involve five steps (uraif), applied across successive experiences:

1. Extraction of information from object states over time (updating/perceptual)
2. Utilizing multiple forms of representation to capture important aspects of change in state-time (representational)
3. Analysis of representations, according to task requirements (analytical):
   3a extraction of the regularity captured by the sequence, or
   3b abstraction of causal relations linking object features to effects, including integration and coordination of operative variables from single observation
4. Further inference, possibly by mental imaging, of a mechanism to link these causal features to the effects (imagery).

5. Evaluation against observed change over time to modify/confirm the enriched segmented representations (feedback)

The data provide clear evidence that the ability to extract information about processes from their spatial and temporal characteristics is established within the elementary years. However, typical development in the elementary age range seems to cover the first three steps. For instance, not all children succeeded even on the flow of liquid task while the ability to coordinate variables was an important interim step. Thus the ability to utilize some analytical and imagery forms in spatial-temporal analysis may be developing slowly. For instance, limited amount of children seemed to manage to employ the fourth step in their causal inference. Most of them still need to construct the actual content of ideas about mechanisms. Although the fifth, feedback, was not directly tested, there are signs that those children who showed advanced causal understanding employed active feedback for their representations.

A key issue in this -failing to apply spatial-temporal analysis to causal processes- may be that space and time are continuous, and extracting temporal information continuously is difficult in particular for the fast processes. By slowing down and segmenting temporal flow, it becomes possible to analyse and reflect on temporally ordered change. This may include construction of images and ideas about transformation – a kind of analysis not captured by spatial tasks, and one which would facilitate analysis of invisible properties. This view fits with ideas that temporal processing may contribute to robust establishment of schematic organization and promotes imagery that would facilitate analysis of invisible properties, such as force or density.
Chapter 5

Overall discussion

Causal reasoning is not limited to the comprehension of distinct causal events, but also requires the grasp of temporally extended processes. The determinants of thinking about such processes are poorly understood, but highly crucial for improving our awareness about the natural world. This thesis revealed various unique findings each pointing to the fact that the competences involved in reasoning about causal processes are quite complex. For instance, although the tasks are presented with the same protocol, most children seemed to find difficult to talk about the kinds of phenomena, and therefore their performances varied for each causal task. Prior knowledge did not play a major role in children’s inferences, but instead more complex patterns needed to be considered to see the holistic picture. An important contributor to thinking about such processes seems to be spatial-temporal analysis. However, this is not a unitary process, some forms of spatial-temporal analysis appear to be more predictive among the other candidates. The predictive power of these forms was consistent across the development, but maturation procedures followed distinct paths. For instance, the analytical form of spatial-temporal analysis seems to be an important driver very early on, as it helps with the segmentation of rapid continuous processes. The ability to follow rapid forms of spatial-temporal qualities seems to develop later, with a reduced predictive power in inferring continuous processes, probably due to its perceptual nature. Verbal ability is another important predictor, but highly susceptible to socioeconomic status. Statistical reasoning has indirect impacts on inferring single instances, highlighting the complex nature of thinking about causal processes. We need to bear in mind that this thesis provides the first systematic investigation that employs process-based causal tasks and examines the function of these candidate abilities. The results need to be discussed in more details as presented under the six subheadings below.

5.1 Children’s explanations progress differently in specific domains

Despite the fact that the topics were within the national curriculum content, children’s explanations differed for specific domains. There were clear differences between sinking, absorption, and solution, in particular with regard to the speed of the processes: required
attention span changed depending on the speed of the processes in the tasks. As discussed before, results here resonate with studies suggesting that slow processes may require longer attention span and may obstruct some children to understand via observation (see e.g. Maurice-Naville and Montangero, 1992; Rieber, 1991). However, the patterns are quite complex, the speed of the process may be only one piece of the puzzle.

Focusing on the scientific discovery process, Klahr and Dunbar (1988) inform us that three categories play role in children’s performance: linguistic knowledge, prior knowledge (e.g. programming knowledge of a computer controlled device), and task specific knowledge, suggesting that children’s performance is neither a mere reflection of the environment nor mental structures. It is more about a joint of the task content and child’s prior experiences. This point of view also considers that knowledge and tasks are not completely distinct, but mutually codetermining factors of the child’s performance (see also Smith et al., 1999; Hewson & Hewson, 1983, for supporting evidence). These components were detectible across the three studies presented in this thesis. For instance, regarding the cognitive factors, only typically developing children involved in all studies, and children’s cognitive task responses (e.g. WASI vocabulary and block design) varied depending on the sample characteristics, but in general exhibited a normal distribution. In parallel with this, almost all children were able to provide some sort of explanation for the causal tasks, though the majority of them considered the basic factors, such as weight, softness, porousness, and therefore their responses ranged in the middle of the scale. Children were not given any clues about the importance of coordinating variables, such as weight-density. There were no practice trials; children did not receive feedback during or after the sessions to minimize the learning effect. Children’s performances were expected to be the function of the cognitive abilities blended with their prior knowledge then. However, it is still too difficult to interpret to what extent children’s explanations was the function of the mere cognitive abilities or prior knowledge. Although the role of prior knowledge was investigated in various ways across three studies, the results consistently showed that neither the role of prior knowledge nor cognitive competences were large in inferring causal phenomena within typical development. There were further factors instead qualifying children’s performances, such as socioeconomic background as indexed by the family income and school catchment, and spatial-temporal analysis independent of other variables. One can suggest that socioeconomic background is an
umbrella term; it may affect various cognitive competences. However, the finding here is unique and introductory. Unique, because the role of socioeconomic status in causal reasoning has not been investigated in great extent. Introductory as because these points plainly merit further investigations across larger samples.

5.2 Can children grasp of causal mechanism from observation?
Research highlights that intervention provides an advantage over observation especially when the focus is learning. However, it is not clear whether intervention overrides observation when the focus is not learning/instruction, but reasoning. A counter-view can suggest that both causal knowledge acquisition and causal reasoning tap into the same cognitive abilities. However, studies warn us to be cautious about that. For instance, Barnett and Ceci (2002); Lecoutre et al. (2016) highlight that there is no clear evidence for the transfer of learning or implementation of prior knowledge to novel situations. Cattell (1963) also proposes that reasoning is the core of fluid intelligence (Gf), while acquired knowledge and skills (e.g. reading, writing, vocabulary) are more relevant to crystallized intelligence (Gc; see also Schneider & McGrew, 2012). Caroll (1993) expands this view by proposing a hierarchical model: the general intelligence factor is placed at the top, Gf and Gc stands separately in level two, with various narrower specific abilities at level one. Nonverbal and spatial abilities are considered as part of Gf, implying less socio-cultural influence on knowledge and intelligence measures. Although this strand of work is beyond the scope of this thesis for now, the message is critical: nonverbal representations of cause-effect relationships take the form of sensory impressions observed mostly from spatial-temporal characteristics of dynamic processes/events. This raises the possibility that reasoning stands as an indicator of ‘potential’ for perceptual and procedural knowledge (e.g. how the cause produces its effect). The procedural part may be relevant to the ability to evaluate the rules/logic, determine the validity of cause-effect relationships, or what information is relevant to causal relationship (e.g. considering invisible intervening factors). Reasoning of course provides benefits for causal learning in a way that it aids the ability to acquire new knowledge, or revise previous causal beliefs. In short, learning stresses memory, reasoning emphasizes ability. This view does not deny the role of observation or intervention for the grasp of causal knowledge, however. The point is that children can causally observe/reason without relying on learning.
On the role of observation, the results here show that some children can grasp causal mechanism by observing a single natural phenomenon. However, the percentage of those who are systematically able to do so across three causal measures is quite small (6.8% for three phenomena; 16.7% for a single phenomenon across three sample N=387). These children typically have near perfect scores in all measures, and their verbal and nonverbal scores are consistently above average. It is plausible to suggest that understanding of causal mechanisms in natural phenomena may be a domain general performance, however, seems like not naturally well developed in the primary school years, probably needs instruction. A short intervention trial tested whether instruction and imaginative games helped children to grasp causal mechanism. This trial showed that children’s awareness of causal mechanism is highly malleable, even if they do not directly manipulate the causes and effects. Given that abstract nature of causal processes may not be responsive to direct interventions, further study can elaborate whether intervention overrides observation when the concern is capturing causal processes rather than contiguous events.

Typically developing children consistently observe and intervene in their environments. But the ability to extract information from their manipulations undergoes development shaped by individual differences. This is the point where individual differences approach delivers an important message: many factors play role conjointly in the maturation of observation of cause-effect relations in the sense that (1) information presentation does matter. Children can benefit seeing contrast items in a simultaneous/successive fashion. (2) Although seeing contrasting items improve children’s inferences, it does not apply to mechanism level thinking, as their mechanism level thinking does not depend on the kinds of demonstrations. (3) The speed of the causal process seems to be another factor (e.g. sinking occurs faster than dissolving) in the sense that (4) children’s ability to use of cognitive resources, such as memory and attention, varies largely. (5) Prior knowledge weakly influences inference and mechanism level responses. Weakly because, the influence of prior knowledge is greater if relevant information presented simultaneously, shortening the time gap between the presentations of the contrast items. Depending on the difficulty of the task, children seem to need some time to organize their thoughts. However, even if they are provided enough space, it does not rule out the influence of
prior knowledge. The three studies employing different methodologies in this thesis showed that predictive power of prior knowledge is quite limited. (6) Across all conditions, only inferential ability consistently associates with mechanism level responses regardless of age, indicating that children those who likely to think about variants beyond the observables, are likely to consider unobservable factors to grasp the processes generating the outcomes the way that they occur.

The question is that “Whether children can benefit from further exercises?” instead. In responding to that, the scientific method allowed children to observe three more objects across three experiments (15 objects in total). Surprisingly, mechanism level responses remained rare, made children to think about three further objects, and plus their correspondences. Alternatively then (7) the picture may be more complex, as (8) the ability to utilise spatial-temporal information seems to require multiple forms as discussed in the fourth chapter (uraif), in particular analytical, imagery and feedback requiring advance cognitive resources beyond memory and attention, (9) corresponding also to the inferential competences employed in statistical thinking.

Further, (10) children’s socioeconomic backgrounds (SES) need to be counted. For instance, none of the children (n=62) from one of the low SES schools in Oxford provided a systematic mechanism level response to the causal tasks. Of course other schools distributed the SES normally, but a careful investigation across the whole sample showed that children who provided systematic understanding of causal phenomena came either from middle or higher SES homes/school catchments. Later informal conversations with children and teachers provided insights into this result. It was stated that parental expectations for their child pursuing a science-related career (teacher perspective), and science related activities in the school (student perspective) are crucial. This indicates that not always within-child factors matter, but also between-child factors have influence on causal thinking.

5.3 How far verbal and nonverbal abilities important for causal thinking?

Given the philosophical and psycholinguistic background require to discussing this question, and the dominant view in the psychology literature propose that language shape
thoughts and therefore causal explanations (see e.g. Piaget, 1959; Vygotsky, 1986), this is a difficult question to answer. For the first time, the thesis investigated the predictive power of verbal, nonverbal and other candidate abilities to assess the extent to which they predicted reasoning about causal processes, but models are restricted to correlational designs. The predictive power of verbal and nonverbal tasks gave mixed messages; the results were dependent on the sample characteristics, in particular SES background. A further investigation combined two samples and employed a quantitative and qualitative analysis to explore the role of verbal and nonverbal abilities in children’s reasoning about causal processes further (Dündar-Coecke & Tolmie, 2019a). This large sample demonstrated a normal distribution regarding the children’s SES background. Results indicated that both verbal and nonverbal ability were unique and major contributors to children’s reasoning of causal processes. However, verbal ability was a stronger predictor of lower levels of causal thinking (factors), and nonverbal ability of higher (variants and unobservables). Although verbal and nonverbal forms of thinking showed widespread dissociations, the effects were not completely independent. However, their overlap in the prediction of reasoning about causal processes reflected their joint influence at lower levels rather than the high. This is consistent with connection between nonverbal and verbal processing being easier when children’s focus was on the static/spatial features of the objects involved in the causal processes.

As children shifted to thinking about dynamic underlying mechanisms, the dissociation became more marked, and the impact of generic verbal ability reduced substantially while nonverbal ability increased. This supports the argument that there is a trade-off in the use of verbal and nonverbal abilities in dynamic causal processes when children have no ready language available to connect perceptions to words.

However, the patterns are not as simple as this. During testing, it was observable that some children, in particular those exhibit good level of mechanism understanding, used more sophisticated language/scientific terms. Looking at children’s causal task responses, in total 101 of them selected randomly in equal numbers from those who had low, middle and high scores on causal measures. Their responses subsampled further for use of scientific vocabulary. This subsample did not differ significantly from the remaining
children on age, parental occupation/education, vocabulary or block design; only
difference was their causal measure scores were marginally higher. Scientific vocabulary
was defined as terms that went beyond the everyday conversation typically used by
children (i.e. specialised terminology capturing aspects of the three phenomena in more
precise fashion). The number of unique scientific terms totalled for each child to give an
overall score, further analyses were run by including children’s scientific vocabulary
scores.

Access to scientific vocabulary assisted performance, and its impact was distinct from
generic vocabulary. Moreover, at the higher levels of causal performance (scores of 3 and
4), explained variance by scientific vocabulary was greater, and its effect overlapped with
nonverbal ability. At lower level (scores of 1 and 2) it correlated with generic vocabulary.
This suggests that scientific vocabulary may grow out of everyday language competence,
but it appears to bridge more to nonverbal knowledge and probably be more relevant to
imagery. However, the measure of scientific vocabulary used here was limited to what
children said or use i.e. it captured performance not competence, so any conclusion must
be provisional. There is substantial need for a standardised measure of comprehension of
scientific vocabulary to support further research on its impact.

This thesis provided a more sensitive index (one-to-one testing rather than group, or
computer-based, or distance testing) of whether children’s home environments had an
influence on these findings. Parental education was a significant factor as one of the most
important socioeconomic parameters. These effects were greater on nonverbal rather than
verbal ability, sharpening the pattern of differential impact of nonverbal ability at higher
levels of reasoning about causal processes.

At the highest levels of causal thinking the effects of nonverbal and spatial-temporal
ability uniquely come to the fore. Their predictive power did not change depending on the
SES factor. However, at the very highest level of causal thinking by parental education
the effects of nonverbal ability waned. This can be thought as children’s focus on the
mechanisms underlying continuous causal processes grew, they drew on nonverbal, in
particular spatial-temporal rather than generic verbal ability. However, the notion of
‘nonverbal’ is a generic term, and its forms need to be explored at both neural and behavioural level.

5.4 On the nature of spatial-temporal analysis

Many qualitative and quantitative attributes of spatial-temporal information influence the way cause-effect relations are encoded, such as topological (i.e. connectedness), metrics (i.e. distance), Euclidian, visual (i.e. color, gloss), and temporal (i.e. movement, succession, growth). Most of these require both spatial and temporal forms to be incorporated into conceptual and schematic representations conjointly, in particular when it comes to extracting causal information the incorporation appears to be crucial.

Looking at behavioral outcomes, the relationship between space and time seems to be asymmetrical. It is not clear whether thinking about time is possible without the involvement of spatial elements, but the data here illustrate that thinking about space without time is possible, though not significantly meaningful for causal inference. Most forms of spatial tasks lack temporal element, and they rely on pictorial/discrete rather than dynamic representations, so they do not provide feedback for future/past outcomes. There are of course studies focusing on the symmetry/asymmetry between space-time representations. Lakoff and Johnson (1999) for instance argue that although we can use temporal words to talk about space, mapping time to space is more difficult than mapping space to time. Similarly, studies with adults (Casasanto & Boroditsky, 2008), and with children (Bottini & Casasanto, 2013) confirm the space-time asymmetry when non-linguistic psychophysical judgments are considered (e.g. participants judged the various spatial lengths that shown on a computerized screen for varying durations. They estimated either duration or spatial length of each line by clicking the mouse). The main message is that disentangling time from space is quite challenging, as various spatial elements readily implicit in temporal tasks, and vice versa. Walsh (2003) also argues that space and time are represented in the brain and mind by a common magnitude system. Taking into account various neuroimaging studies, he proposes a model relying on overlapping cortical areas for spatial and temporal processing. Based on a behavioral approach this thesis revealed the lack of predictive power of spatial tasks in continuous causal processes, instead emphasized the predictive aspects of spatial-temporal analysis.
The results here seemingly communicate with Bottini and Casasanto’s findings in the sense that both studies focus on the psychophysical processes. Their findings seem to be consistent on the role of spatial information in temporal thinking. However, this thesis provided more conclusive arguments supported by the data on children’s thinking about time based on spatial representations, and their conjoint affects on causal thinking.

The three spatial-temporal tasks (liquid flow, speed, DTV) required participants to integrate temporal components such as distance, duration, and speed in dynamic fashion beyond mental transformations of static visual stimuli. The amount of information changed for ‘two/three dimensional spatial’ versus ‘four-dimensional spatial-temporal’ displays, and therefore these types necessitated different amount of cognitive loads. Liquid flow task was computed sequentially therefore required participants to put the sequences in temporal order by organising of their representations. With the speed children needed update online approximations of the three moving objects. Velocity (in DTV) required imagery. All together, several forms of spatial-temporal analysis were trialed. Some of these forms showed similar patterns (e.g. DTV total and liquid flow, but the components of DTV especially time and velocity differed). It was hard to interpret children’s performance on speed task at the beginning. Further, comparing between the high versus low achievers’ performances on speed, the data shoed that children who provided consistent mechanism level responses did well with the speed while the low achievers struggled with the task most. In the time tasks, participants did not see the entire run, but instead they needed to estimate the future outcome. For some children, seeing the current and estimating the future position of one moving item (indexed by time in DTV task) was more easier than comparing the speed and distance of three moving items (indexed by speed tasks); and the former was also a better predictor. This was not the case for older children and adults (cf. Dündar-Coecke, 2019) however, as they were capable of utilising spatial-temporal properties of three moving objects either in virtual or actual environment. It is therefore highly likely that reasoning about time (1) in segmented, (2) online/short time, and (3) imagery settings rely on distinct cognitive processes possibly develop with different rate.
On the symmetry/asymmetry, although it is highly difficult to understand from behavioral designs, the developmental perspective appears to capture some initial signs about the role of spatial-temporal processing in inferring causal relations. This thesis suggest that spatial-temporal qualities might be administrated over the steps involving updating, representing, analyzing, imagining, and feedback, and there may be a cyclic loading among them. Organization of these to (re)construct dynamic representations might be crucial for causal inference, with the steps allowing other kinds of reasoning modes to be involved, in particular from the third stage (analysis) and onwards. However, this point merits further behavioral and neuroimaging investigations.

The role of temporal element is critical here. It is shown by the previous work (see Hoerl and McCormack, 2018; and McCormack and Hoerl, 2017 for a view) that temporal updating and engaging in temporal reasoning require different cognitive systems, subject to the development, while the latter corresponding to mechanisms in which operates with a model of the world. In this thesis one of the spatial-temporal tasks –speed- primarily relied on temporal updating. Others –such as liquid flow- needed keeping track of time and also maintenance of representations of past, current, and future elements beyond updating. Consistently, this was more predictive of causal processes as well as mechanism level thinking. Moreover, the low correlations between ‘spatial’ and ’spatial-temporal’, and the lack of predictive power of the spatial tasks provide further evidence on the importance of the spatial and temporal convergence.

For now, this thesis produces two main arguments; the first proposes that spatial-temporal analysis might be a unique form of nonverbal intelligence as discussed below. The second disagrees with the views suggesting a competition between spatial-temporal and other forms of thinking (e.g. Harris et al, 1996). Any other kinds of thinking, including counterfactual, need prior input for a counter view. Logically, input precedes comparison, and this can be done by spatial-temporal analysis. The ability to extract information from space and time, which can be utilized in any kind of reasoning ranging from analogical to counterfactual, seems to be a general function of intelligence then.
This thesis showed that the forms of spatial-temporal analysis could explain up to 15.6% (5.1% speed; 7.7% liquid flow, 2.8% DTV) variance of thinking about causal processes over and above other candidate abilities. However, our lack of understanding remains whether spatial-temporal tasks require more perceptual or higher inferential resources. Behavioural studies have provided initial signs for what kind of cognitive processes may be involved in this particular thinking. But, neural analyses are needed to elaborate which forms of spatial-temporal analysis are most important for causal inference across development and what are their correspondences with other competences.

Considering the possibility of its domain-general characteristics it is plausible to argue that spatial-temporal analysis may be another form of nonverbal intelligence. The data have consistently showed that there are multimodal cognitive styles in the context of causal thinking; neither verbal nor nonverbal abilities are unitary. For instance, although mapped onto verbal domain, use of scientific vocabulary is a unique precursor lying in the interface of verbal-nonverbal abilities, suggesting that access to more specialized scientific vocabulary may assist connection between nonverbal representations and generic language. Moreover, while spatial forms share their variances with verbal, none of the spatial-temporal measures did so. As it was discussed earlier, nonverbal intelligence is poorly understood, its forms need to be explored at both behavioral and neural level. Unfortunately, our understanding is limited to the initial studies presented in this thesis.

5.5 Can children understand causal mechanism from single observations?

In this thesis studies presented with a design where causal relations needed to be inferred from single instances. In the end, spatial-temporal, non-verbal, and verbal abilities appeared to be the strongest predictors. Statistical thinking also played indirect role in interpreting even a single causal phenomenon, supporting the view that cause-effect relations can be judged not only from regularities/statistical data, but also that of single demonstrations of temporally extended dynamic processes.

It appears that probability, as a tool, provides children with a sense of awareness of uncertainty and unseen factors that allow exceptions. Although the methodology did not
require children to explain the way they link probability and the three causal experiments, informal conversations with some of high achievers provided insights into this finding: even for single instances, they consider the possibility of other factors which are not readily available. Plus, most children found it crucial to see two objects falling/rising/dissolving differently in the same environment, which can be thought as a primitive form of covariation detection. Together with spatial-temporal thinking, those two competences seem to provide a fundamental to thinking about a single causal phenomenon. As the path model illustrated, an extended structure, rather than the diverse effects, suited to this explanation. A holistic perspective combining these three competences may suggest an innovative basis for early scientific literacy: analysing spatial-temporal characteristics, comparing similarities/dissimilarities for each factor, and allowing exceptions in conclusions. Further interventions can trial the impact of this outcome.

5.6 The meaning of these outcomes for education

At present, science lessons typically focus on the ‘what’ rather than the ‘why’ or ‘how’, and do not actively support causal thinking, neither spatial-temporal analysis. Understanding of causal mechanisms has largely been ignored in past work on scientific reasoning, with studies typically assessing multiple aspects of children’s knowledge or focusing on their explanations without differentiating between accounts of factors, variables and mechanisms. Across the whole sample, very limited amount of children were able to explicitly report mechanisms for causal processes. This percentage was much lower for young children, and although mechanism awareness was detectible in their thinking, this varied highly depending on domain general abilities (e.g. verbal, nonverbal) and socioeconomic background. Further investigations indicated that limited awareness of mechanism has measurable negative effects on school science attainment. For instance, children’s mechanism inference correlated with their performance on 2011 TIMSS items (Trends in International Mathematics and Science Study), suggesting that understanding causal mechanism might be the key for problem solving skills in other areas.
Further investigations can hypothesize that for school science, activities need to place much greater emphasis on promoting spatial-temporal and nonverbal awareness of causal processes, even connecting these directly to appropriate scientific vocabulary. A classroom exercise has trialled this approach (Dündar-Coecke & Tolmie, 2019b). The impact of this short-term intervention designed to promote spatial-temporal thinking with regard to one such process, sinking. The exercises produced substantial improvements in children’s performance, regardless of age. The results showed increase in the accuracy of children’s initial and final predictions broken down by class. There were good grounds for concluding that the intervention had widespread immediate effects on children’s understanding. Given their socioeconomic background – the majority of children came from low SES families – this effect was promising.

The take-home message from this trial is that the ability to analyze spatial-temporal information appears to be malleable even in a short period of time; children can improve their understanding about causal processes; how to become aware of seen and unseen dimensions, and use them in their causal analysis. Therefore, analysis of spatial-temporal information to actual imaging of causal mechanisms may dynamically tie these elements together – the key step that the majority of primary age children seem unable to make these within current science teaching.
References


Dowe, P. (2000). *Physical causation* (Cambridge studies in probability, induction and


NY: Guilford.


Kant, I. (1783). *Prolegomena to every future metaphysics that may be presented as a science*. New York: Modern Library.


Keil, F. (1979). The development of the young child’s ability to anticipate the outcomes of simple causal events. *Child Development*, 50, 455-462.


The Official Journal of the National Association for Research in Science Teaching, 43, 320-347.


45-61.


Appendix 1

Measures

Virtual speed task 1 (VS1). The initial version of this task developed and used in the chapter three with 17 trials. On a Macintosh laptop (resolution 1440 x 900 pixels) participants saw computer animations of three bunnies (red, yellow, black) racing towards a carrot from different start positions at different speeds (Figure 22a), with the animation stopping before they reached it. Children judged which bunny would arrive at the carrot first. The task began with two practice items, followed by 13 trials gradually increasing in difficulty: the stop time reduced, from 4 to 2 seconds, as did the difference between the three bunnies in start point and relative speed, making differences in arrival time harder to distinguish, and the period available within which to track the differences shorter. The number of correct responses was recorded (0-13). At the end, participants were asked to judge the difficulty level of the task, and make commend if they had any.

![Figure 22a](image1.png)

(a) VS1

![Figure 22b](image2.png)

(b) VS2

Figure 22. Example configurations of bunnies at the start of

(a) VS1, and (b) VS2 tasks

Virtual speed task 2 (VS2). This task was exactly the same as the previous speed task with one difference: rather than three bunnies, participants saw computer animations of two bunnies (black and yellow to ensure the visual discernibility) racing toward the target item from different start positions at different speeds (Figure 22b). The original speed task used in chapter three ranged in difficulty, positively skewed with a long tail
in young children’s responses, indicating that majority of primary school children (5-to-11-year-olds) in particular younger age found the task taxing. Further analyses showed that the majority of children struggled with 8 trials consistently. These 8 trials were chosen to elaborate whether the intensity of the information load played role in this. The 8 trials were demonstrated with less intensity by cancelling one bunny located in the middle in all the trials to keep the distance as constant (in some trials the colour of one of the bunnies needed to be changed to keep the consistency see e.g. Figure 22b). Therefore the distance between the two bunnies widened. The number of correct responses was ranged between 0 and 8. Similarly, participants were asked to judge the difficulty level of the task, and make commend if they had any.

Actual speed task (AS). This task was adopted from Piaget (1969/2006) who demonstrated that primitive/early understanding of space and time was highly dependent on duration-distance judgments. Participants compared three clockwork toys in each trial (e.g. a snake, a crocodile, and a doggy), differing on speed, duration and distance towards an end point. Participants were shown the half of the run, the other half was hidden by using a cardboard tunnel. Durations of travels varied either 8 or 10 seconds, but the visible part were ranged from 4 to 5 seconds with varying distance. In total nine clockwork toys were used, two of them were replaced in each trial to avoid conditioning effect. Once the winding keys were set up, the three objects were put behind a cardboard wall with varying distance/angle from the tunnel, therefore the race was started equally. Participants needed to guess the winner from the distance to the end point and velocity of the object. The three objects traveled either (a) different distance with the same time, or (c) same distance with different time. Each time only one item travelled the longer distance with highest speed. The number of correct responses was ranged between 0 and 5 over five trials.

Results
Analyses utilised data from all 52 participants who completed testing. The observed power for ANOVA was 0.90, for regression analyses it was 0.98. The means for each age group illustrated that young children did not perform well (proportionate to the maximum) on VS1, but instead they relatively performed well on VS2, and well on AS
tasks (Table 28). A one-way ANOVA showed significant age-related progression on VS1 and VS2 performance, however, for the VS1 differences between the groups was highly significant \( F(2,51)=35.525, p<.001 \), partial eta squared=.418, Welch and Brown-Forsythe robust tests were also highly significant \( p<.001 \), indicating later growth for responses in this task.

Table 28. Mean score (sd) by age group on VS1 (max=13), VS2 (max=8), and AS (max=3)

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Nursery</th>
<th>Reception</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>VS1</td>
<td>5.3 (3.2)</td>
<td>8.7 (2.8)</td>
<td>12.9 (0.4)</td>
</tr>
<tr>
<td>VS2</td>
<td>6.3 (1.8)</td>
<td>7.4 (0.9)</td>
<td>8.0 (0.0)</td>
</tr>
<tr>
<td>AS</td>
<td>2.8 (0.4)</td>
<td>3.0 (0.0)</td>
<td>3.0 (0.0)</td>
</tr>
</tbody>
</table>

Zero-order correlations showed that there were no significant associations between the tasks and either gender or their socioeconomic status (Table 29).

Table 29. Zero-order and partial correlations

<table>
<thead>
<tr>
<th></th>
<th>VS1</th>
<th>VS2</th>
<th>AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.767***</td>
<td>.516***</td>
<td>.305***</td>
</tr>
<tr>
<td>VS1</td>
<td>1</td>
<td>.733***</td>
<td>.364**</td>
</tr>
<tr>
<td>VS2</td>
<td>.613***</td>
<td>1</td>
<td>.102</td>
</tr>
<tr>
<td>AS</td>
<td>.214</td>
<td>-.068</td>
<td>1</td>
</tr>
</tbody>
</table>

Zero-order correlations above diagonal, \( N=52 \); partial correlations below diagonal; *\( p<.05 \), **\( p<.01 \), ***\( p<.001 \)

Age highly correlated with VS1 \( (=r.767, p<.001) \) and VS2 \( (r=.516, p<.001) \) respectively, confirming the ANOVA results; VS1 highly correlated with VS2 \( (r=.733, p<.001) \) and moderately with AS \( (.364, p<.05) \). However, VS2 did not associate with AS \( (r=.102.\)
p>.05). When controlling for age, the patterns remained similar with a slight reduction in correlations $r(\text{partial})=.613$, $p<.001$; and $.214$, $p<.01$, suggesting that age was not the only factors, but the characteristics of the tasks did also matter.

Further regression analyses showed that age explained the majority of variance (58.8%) in VS1 performance ($\beta=.767$, $p<.001$); 26.6% of variance in VS2 ($\beta=.516$, $p<.001$), and nearly 9% in AS performance ($\beta=.305$, $p<.05$), either SES or gender was significant, confirming the high correlations. Overall, age effect was highest on VS1 and slightly less on VS2, non significant for AS, suggesting that age effect was highly significant in virtual task, but it disappeared in the actual speed task.