Causal Discovery: How infants and adults represent complex causal relations

by

Deon T. Benton

A Dissertation Presented in Partial Fulfillment

of Requirements for the Degree of

Doctor of Philosophy

Graduate Supervisory Committee

Dr. David H. Rakison, Chair

Dr. Anna Fisher

Dr. David Danks

Department of Psychology

Carnegie Mellon University

© Deon T. Benton

All Rights Reserved, 2018

Abstract

The ability to reason about causal events in the world is fundamental to cognition. Despite the importance of this ability, little is known about how infants and adults represent causal events, what structure or form those representations take, and whether such events are represented as causal graphical models or learned associations between low-level perceptual features. It is also unknown whether infants and adults are sensitive to the Markov condition in causal-graphical models—in which a first object or entity is independent of other objects or entities conditional on its parents—and can form Markov-equivalence classes; that is, it remains unknown whether infants and adults treat equivalently causal events that instantiate identical conditional independence relations. The aim of this dissertation is to examine these issues in four experiments with adults and one experiment with infants between 18 and 22 months of age using four-object, three-chain causal launching-event sequences. The experiments reported here were designed specifically to examine when, whether, and to what extent infants and adults process four-object, three-chain launching-event sequences in terms of low-level perceptual cues that are readily available in the perceptual input or in terms of the Markov condition and Markov-equivalence classes. Launching-event sequences were used because they are the simplest events in which to instantiate, and to observe, cause-and-effect relations. Four objects were used because it is possible to examine infants' and adults' sensitivity to multiple conditional independence relations in four-chain, but neither in two-chain nor three-chain, launching sequences and due to the relatively large number of possible combinations in which to arrange the four objects permit (i.e., 24). The results of Experiments 1 to 4 revealed that, although adults showed restricted sensitivity to the Markov condition and Markov-equivalence classes, this sensitivity was disrupted by low-level perceptual manipulations to the launching-event sequences. The results of Experiments 1 to 4 revealed additionally that adults showed a greater tendency to respond to the four-object launching sequences in terms of low-level perceptual cues such as the spatial relation between objects than in terms of the Markov condition or Markov-equivalence classes in which two or more CGMs instantiate identical probability information. In contrast to the results of Experiments 1 to 4, the results of Experiment 5 revealed that it was neither the case that infants between 18 and 22 months of age processed the events in terms of the Markov condition and Markov-equivalence classes nor was it the case that they processed the events it terms of low-level perceptual features. The results indicate that infants do not represent causal events as CGMs and that adults show a limited ability to do so that facilitated in some contexts but not in others. Taken together, these results cast serious doubts on theories that claim that humans represent causal events as CGMs and suggest that it is more likely that such events—especially those used here—are represented in terms of readily available low-level perceptual information.

Acknowledgments

[Insert text here]

TABLE OF CONTENT

[1 Introduction 1](#_Toc522727389)

[1.1 Overview of the present dissertation 1](#_Toc522727390)

[1.2 Defining causal perception and causal reasoning 3](#_Toc522727391)

[1.3 Summary 8](#_Toc522727392)

[2 Studies on causal perception and corresponding theories 10](#_Toc522727393)

[2.1 The launching-event paradigm 10](#_Toc522727394)

[2.2 Domain-specific theories of causal perception 11](#_Toc522727395)

[2.3 Domain-general theories of causal perception 18](#_Toc522727396)

[2.4 Summary 23](#_Toc522727397)

[3 Studies on causal reasoning and corresponding theories 24](#_Toc522727398)

[3.1 Associative models of causal reasoning: The Rescorla-Wagner model 24](#_Toc522727399)

[3.2 Associative models of causal reasoning: The modified Rescorla-Wagner model 27](#_Toc522727400)

[3.3 Contrast-contingency models of causal reasoning 28](#_Toc522727401)

[3.4 The Power PC Theory 31](#_Toc522727402)

[3.5 The Bayesian Inference perspective 33](#_Toc522727403)

[3.6 Causal Graphical Models 37](#_Toc522727404)

[3.7 Current Experiments 46](#_Toc522727405)

[4 Experiment 1 49](#_Toc522727406)

[4.1 Methods 50](#_Toc522727407)

[4.2 Predictions 54](#_Toc522727408)

[4.3 Results 56](#_Toc522727409)

[4.4 Discussion 63](#_Toc522727410)

[5 Experiment 2 66](#_Toc522727411)

[5.1 Methods 67](#_Toc522727412)

[5.2 Predictions 68](#_Toc522727413)

[5.3 Results 70](#_Toc522727414)

[5.4 Discussion 75](#_Toc522727415)

[6 Experiment 3 78](#_Toc522727416)

[6.1 Methods 78](#_Toc522727417)

[6.2 Predictions 81](#_Toc522727418)

[6.3 Results 82](#_Toc522727419)

[6.4 Discussion 87](#_Toc522727420)

[7 Experiment 4 89](#_Toc522727421)

[7.1 Methods 89](#_Toc522727422)

[7.2 Predictions 91](#_Toc522727423)

[7.3 Results 93](#_Toc522727424)

[7.4 Discussion 98](#_Toc522727425)

[8 Experiment 5 100](#_Toc522727426)

[8.1 Method 100](#_Toc522727427)

[8.2 Results 103](#_Toc522727428)

[8.3 Discussion 115](#_Toc522727429)

[9 General Discussion 116](#_Toc522727430)

[9.1 Summary of Adult Experiments 118](#_Toc522727431)

[9.2 Summary of Infant Experiments 124](#_Toc522727432)

[9.3 Conclusion 134](#_Toc522727433)

**LIST OF FIGURES**

*Figure 1A-C*. Examples of a (A) standard launching event sequence that is composed of two objects, (B) a single causal chain that is composed of three objects, and (C) two causal chains each of which are composed of three objects38  
*Figure 2A-D*. Examples of the training event and test events used in Experiment 1. Subjects were shown the training event three times from left-to-right and three times from right-to-left. The position of the red and blue objects was counterbalanced across subjects. Subjects then saw the four test events, and the test events were shown in counterbalanced order across subjects53   
*Figure 3*. Predicted pattern of responding to the test events in Experiment 1 based on whether subjects responded to the test events based on independence-relation information, the Markov condition, temporal-precedence information, or the novelty of the test events relative to the training event56   
*Figure 4*. Subjects’ perceptual and causal ratings of the four test events in Experiment 1. The black bars correspond to subjects’ perceptual ratings of the test events and the gray bars correspond to subjects’ causal ratings of the test events57   
*Figure 5*. Each of the 64 subjects’ causal and perceptual ratings of the four test events in Experiment 161  
*Figure 6*. Predicted pattern of responding to the test events in Experiment 2 based on whether subjects responded to the test events based on differences relative to the training event in six features70  
*Figure 7*. Subjects’ perceptual and causal ratings of the four test events in Experiment 2. The black bars correspond to subjects’ perceptual ratings of the test events and the gray bars correspond to subjects’ causal ratings of the test events73  
*Figure 8*. Each of the 64 subjects’ causal and perceptual ratings of the four test events in Experiment 274  
*Figure 9*. Predicted pattern of responding to the test events in Experiment 3 based on differences relative to the training phase in the causal role of the green object, spatial information of each of the four test event sequences (i.e., whether a spatial gap was included), the Markov condition, or the novelty of entire test event sequences82  
*Figure 10*. Subjects’ perceptual and causal ratings of the four test events in Experiment 3. The black bars correspond to subjects’ perceptual ratings of the test events and the gray bars correspond to subjects’ causal ratings of the test events85  
*Figure 11*. Each of the 32 subjects’ causal and perceptual ratings of the four test events in Experiment 386  
*Figure 12*. Example of the training event and test events used in Experiment 491   
*Figure 13*. Predicted pattern of responding to the test events in Experiment 4 based on whether subjects responded to the test events based differences relative to the training phase in the causal roles of the four objects, color of the four objects, novelty of an entire test event sequences, or familiarity of an entire test event sequences92   
*Figure 14*. Subjects’ perceptual and causal ratings of the four test events in Experiment 4. The black bars correspond to subjects’ perceptual ratings of the test events and the gray bars correspond to subjects’ causal ratings of the test events96  
*Figure 15*. Each of the 32 subjects’ causal and perceptual ratings of the four test events in Experiment 497

*Figure 16*. Infants’ mean looking times to the four test events and one posttest event in Experiment 5107

*Figure 17*. Looking times to the four test events for the “younger” and “older” infants109

*Figure 18*. Looking times to the four test events for the 13 “younger” infants111

*Figure 19*. Looking times to the four test events for the 12 “older” infants112

*Figure 20.* Looking times to the four test events for the “fast” and “slow” habituators113

*Figure 21.* Looking times to the four test events for the 12 “slow” habituators infants114

*Figure 22.* Looking times to the four test events for the 13 “fast” habituators infants115

# Introduction

Of all the abilities that enable humans to understand how the world works, perhaps few are as fundamental as causal perception and causal reasoning. These abilities are important because they allow humans not only to perceive and encode cause-and-effect relations in the world but to reason about the effects of interventions on those relationships as well as to reason about the effects of interventions that were not explicitly undertaken. Successful understanding of how light switches operate, for example, requires that reasoners first recognize that a causal relation exists between the light switch and its source in the first place; that is, it requires that learners first perceive the relation as a causal one in which the light switch is seen as the cause of the light source's illumination. Second, it requires that reasoners be able either to generate the effect or to extinguish it by manipulating the position of the light switch itself (e.g., flipping the switch up) such as by flipping the switch in one direction to produce light illumination but in a different direction to prevent light illumination. Finally, a successful understanding requires that learners be able to reason counterfactually about the relation between causes and effects; that is, a complete causal understanding of this simple light system requires that learners be able to reason about what would have happened if an alternative action had been undertaken such as recognizing that the light bulb would not have illuminated if the switch were flipped in the opposite direction.

## Overview of the present dissertation

Research on this topic with young infants has revealed that the ability to perceive cause-and-effect relations emerges between 4½ and 10 months of age (e.g. Leslie & Keeble, 1987; Oakes & Cohen, 1990; Rakison & Krogh, 2012), but that the ability to engage in more explicit forms of reasoning about causes—such as that described above—is thought to emerge later in development between 18 months and 5 years of age (e.g., Gopnik et al., 2004). Despite the impressive body of work that has examined the development of causal perception and causal reasoning in infants and young children, little is known about: (a) how, whether, and to what extent, causal perception and causal reasoning are related, (b) whether early causal perception underpins later causal reasoning, and (c) whether and to what extent both abilities are subserved by the same or different underlying learning mechanisms. It is as-yet unresolved, for example, whether causal perception and causal reasoning are underpinned by domain-specific learning mechanisms, modules, or skeletal systems that constrain and facilitate causal learning or whether, instead, both abilities are underpinned by domain-general learning mechanisms such as associative learning that operate over structured input and that support learning across a range of content domains.

Comparatively little is also known about: (d) how humans represent causal events, (e) what the structure of those representations are, including what learning mechanism gives rise to these representations, and (f) how this representation changes between infancy and adulthood. The projects discussed in this dissertation explore these latter issues; namely, the experiments reported here use four-object launching sequences to examine how infants (Experiment 1) between the ages of 18 and 22 months and adults (Experiments 1 to 4) represent causal events, whether those representations are structured as causal graphical models or learned associations among and between low-level perceptual cues and features, and how and whether causal representations change between late infancy and adulthood. In terms of the structure of causal representations, the experiments reported here explore whether infants and adults are sensitive to and can encode the Markov condition—which I describe in detail in Section 3.6—in four-object launching sequences and, to the extent that infants and adults can encode the Markov condition, whether (1) both groups respond equivalently to test sequences that preserve the conditional independence relations expressed by the Markov condition but differently to sequences that violate it (Experiment 1 and 4) or, instead, (2) based on lower-level perceptual cues such as the spatial relations between the elements in the sequences (Experiments 2 and 3), and (3) whether responses based on either of these potential cues changes between late infancy and adulthood. If infants and/or adults are shown to be sensitive to and can encode the Markov condition, an additional experiment with adults examines over what variables adults encode the Markov condition (Experiment 4). The experiments reported here are novel because they examine open and as-yet explored questions that pertain to how infants and adults represent causal events, whether those are represented as causal graphical models or learned associations, and how causal representations change between infancy and adulthood. Second, the experiments reported here are novel because they use complex four-object, three-chain launching event sequences that have not been used in previous research to explore novel questions that pertain to how infants and adults process and represent complex causal events and to distinguish between competing theoretical accounts. The experiments reported here are also unique because they attempt to bridge the gap between causal perception and causal reasoning: they were designed to distinguish between, and test the viability of, competing theoretical accounts of causal perception and causal reasoning in the context of putative launching-event sequences.

## Defining causal perception and causal reasoning

Despite the fact that there is an extensive body of research that examines causal perception and causal reasoning in young infants and children, there have been relatively few attempts to define them (though see Danks, 2009 for one attempt) . One reason for this is that researchers have disagreed considerably on the structure of the representation that underlies both abilities and whether both abilities are underpinned by the same or different mechanism. For example, there is little consensus among developmentalists about whether causality in simple launching events—in which one object causes another object to move—is directly perceived (*e.g.*, Leslie, 1994, 1995; Michotte, 1946; White, 1988) or is inferred based on low-level perceptual cues such as spatial and temporal contiguity (*e.g.*, Hume, 1748; Oakes & Cohen, 1990). Likewise, it is unclear whether domain-general or domain-specific mechanisms underpin causal perception and causal reasoning; that is, it is an open question whether specialized mechanisms—that may emerge from birth or shortly thereafter—subserve causal perception and causal reasoning or whether such perception and reasoning is underpinned by general mechanisms that support learning in a range of content areas. For example, proponents of the domain-specific perspective argue that the ability to reason about causal events is underpinned by specialized mechanisms and systems that support causal perception and causal reasoning (*e.g.*, Gopnik, 2003; Gopnik & Wellman, 1994; Leslie, 1994, 1995; Mandler, 1992). In contrast, proponents of the domain-general view—where the same all-purpose mechanism that governs causal learning also governs learning in other domains (e.g., language)—argue that infants, children, and adults learn to perceive and, later, reason about causal events based on the predictive statistics in causal events (i.e., the correlations between presumed causes and effects).

Given this lack of consensus on the definitions of causal perception and causal reasoning, I define causal perception narrowly and in the context of launching-event sequences. I restrict my definition of causal perception to launching sequences because (1) most, if not all, of the studies on causal perception use launching event sequences, (2) launching event sequences are perhaps the clearest and simplest examples of cause-and-effect relations, which have made them especially suitable for testing infants' early causal perception abilities, and (3) launching event sequences can be interpreted without recourse to predicate logic or reasoning. Causal perception can therefore be defined in terms of two abilities. First, it requires the ability to distinguish between causal launching events—that is, events in which the movement of the second object is spatially and temporally contiguous with that of a first object—and non-causal launching events—that is, events that lack either spatial contiguity, temporal contiguity, or both (Oakes & Cohen, 1990; Leslie, 1984). This means that learners who respond to launching events on the basis of their causality should respond equivalently to non-causal launching sequences—despite the fact that any two non-causal sequences may differ (*e.g.*, delayed launching vs. no-collision; discussed below)—but differently to a causal sequence in which a first object causes a second object to move immediately and upon physical contact from a first object.

Second, causal perception must involve the ability to associate specific objects with specific causal roles. Thus, causal perception involves both the ability to distinguish causal from non-causal events but also the ability to associate particular objects in those sequences with the role of agent and other objects with the role of patient or recipient. It is important to note that because the only measure used to assess causal perception in infants is differentially looking (i.e., infants are not required to make explicit decisions about the events) and because causal perception, but neither causal judgements and reasoning, depends on lower-level perceptual cues (*e.g.*, Schlottman & Shanks, 1992), causal perception is thought generally to be an implicit process. In sum, causal perception can be defined as the implicit ability (1) to distinguish causal from non-causal launching events and (2) to associate particular objects with specific causal roles within these events.

Similar to causal perception, causal reasoning can be defined narrowly to refer to children's ability to reason about (mostly physical) causal events in which an object (or series of objects) causes a change in state in another object or series of objects through contact or no contact (*e.g.*, psychological causality or contact at a distance) (though see Bullock, Gelman, & Baillargeon, 1982 for a review of other studies that use the term causal reasoning). I restrict my definition of causal reasoning to physical events because the principal claim of which the experiments described here were designed to test the viability—that is, humans represent causal events as causal graphical models—was first introduced in the context of physical causal events; that is, the notion that humans represent causal events as causal graphical models and use Bayesian inference to choose the most likely model was introduced in the context of blicket-detector experiments in which children are asked to determine which objects are blickets and which are not based on different patterns of activation (described in more detail below).

Note that although I restrict my definition of causal reasoning to physical causal events, the definition I offer nonetheless incorporates several key features that (a) may factor importantly in other forms of causal reasoning such as psychological causal reasoning (*e.g.*, Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003) and that (b) have, until now, been discussed in isolation. In terms of the latter of these two points, no extant definition of causal reasoning attempts to define causal reasoning in terms of multiple features and interrelated abilities but rather tends to focus on just one of many key features and interrelated abilities. These features, which have tended to be discussed in isolation, include the ability to encode causal relations among and between objects and entities through observational learning and based on low-level perceptual cues such as temporal information (*e.g.*, Hagmayer & Waldmann, 2002), to associate particular objects or entities with particular causal roles (*e.g.*, Cohen, Amsel, Redford, & Casasola, 1998), to intervene on those relations to generate novel effects (*e.g.*, Gopnik et al., 2004), and the ability to engage in counterfactual reasoning—that is, to know and predict what would have happened if a different action had been executed (*i.e.*, Harris & German, 1996; Lewis, 1973, 1986; Sobel, 2004).

Despite the fact that associative learning, interventional learning, observational learning, and counterfactual reasoning have all, individually, been suggested to be key features of physical causal reasoning (and features that may impact other kinds of reasoning such as psychological reasoning; e.g., Steyvers, Tenenbaum, Wagenmakers, & Blum, 2002), there is no extant definition that discusses causal reasoning in terms of all four features. I therefore define physical causal reasoning in terms of these four key abilities. In particular, a critical first step of successful causal reasoning, similar to causal perception, is the ability to associate specific objects or events with the role of cause or agent and other objects or events with that of effect or recipient such as when a child in a blicket-detector design forms an association between an object and the detector’s activation. Second, successful causal reasoning requires that causal learners not only associate cause and effects but encode those causal relations through observations to generate predictions. Third, once learners form the appropriate association between potential physical causes and effects and encode that relation through observation, they must be able to intervene on those relations to generate novel effects. Finally, given that much of causal reasoning requires the ability to reason about events that did not occur but nonetheless could have, causal reasoning requires that causal learners be able to generate counterfactual inferences based on observed causal data.

In short, physical causal reasoning can be defined as the ability (1) to associate physical causes and effects in the world, (2) to observe and encode those relations to generate predictions, (3) to intervene on those relations to bring about different effects, and (4) to reason about counterfactual events. A corollary of this definition is that individuals who espouse only a subset of these abilities would not be considered mature causal reasoners, whereas individuals who possess all four abilities would be considered mature reasoners. This way of defining physical causal reasoning suggests a causal processing gradient, whereby individuals who do not possess any of the abilities would fall along one end of the gradient, individuals who possess all four abilities would fall along the other end of the gradient, and individuals who possess only a subset of the abilities would fall somewhere along the gradient in between both ends. Note that although others may disagree with these definitions or have definitions of their own, to my knowledge this is one of the first times that complete and explicit definitions of causal perception and causal reasoning have been provided.

## Summary

The primary aim of this section was to provide a coherent and complete definition of causal perception and causal reasoning. This goal was motivated by the fact that no complete definition exists for either ability. Thus, I defined causal perception as the ability implicitly (i.e., without effort or explicit, verbal reasoning) to distinguish causal events from non-causal events and associate certain objects within these events with the role of agent and other objects with the role of recipient. I defined causal reasoning as the ability to observe causal relations in the world and make predictions based on them, intervene on those relations to bring about new effects, and reason counterfactually about those relations.

The goal for the remainder of the dissertation will be to discuss existing research on causal perception and causal reasoning and the theories that have been developed to explain the mechanism or mechanisms that underpin both abilities. Of these accounts, the account that is most relevant to the present experiments is the Bayesian inference account of human causal learning (e.g., Gopnik et al., 2004). I focus on this account specifically for two reasons. First, many believe—based on recent evidence in infants (e.g., Sobel & Kirkham, 2006, 2007), children (e.g., Sobel, Tenenbaum, & Gopnik, 2004), adults (Griffiths, Sobel, Tenenbaum, & Gopnik, 2011), and even tentative evidence in non-human animals (e.g., Povinelli, 2000; Taylor et al., 2014; Tomasello & Call, 1997)—that human and non-human causal reasoning is best explained within the framework of this account. Second, this account makes several assumptions about human causal inference and about the nature of causal representations that are as-yet untested such as that: humans use Bayes' rule to choose among competing causal structures; that humans represent all possible structures; that the likelihood of causal structures is weighted by prior probability information; and that humans represent causal events as causal graphical models and are sensitive to and can encode conditional independence information in these structures according to the faithfulness and causal Markov assumptions.

The experiments discussed in this dissertation focus on this latter assumption; that is, the experiments described here are designed to examine (a) whether 18- to 22-month-old infants and adults represent causal events as causal-graphical models, (b) whether they can encode the Markov condition in these structures (i.e., the conditional independencies in graph structures), or (c) whether, instead, they respond to causal events on the basis of lower-level spatiotemporal cues. It is worth noting that Experiments 1-5 were not devised specifically to examine whether humans use Bayes' rule to choose the most likely structure. This is because a necessary condition for using Bayes' rule to choose the most likely structure is that learners represent causal structures as causal graphical models over which Bayes' rule can operate. Thus, before it can be determined whether, and to what extent, and under what conditions human learners use Bayes' rule to distinguish competing causal graph structures, it is necessary first to establish that learners can represent causal events as causal graphical models in the first place and are sensitive to and can encode the causal Markov condition within these structures.

# Studies on causal perception and corresponding theories

## The launching-event paradigm

In the previous section, we defined causal perception and causal reasoning. These definitions notwithstanding, there is considerable debate about the nature of the representations that underlie causal perception, whether causal perception is an ability that undergoes developmental change, and whether causal perception in launching-event sequences is underpinned by domain-general mechanisms or specialized domain-specific mechanisms (Leslie, 1995; Oakes & Cohen, 1995). Given the trenchant nature of this debate, my aim in the next two sections is to outline the different domain-general and domain-specific theories that pertain to the perception of causality. I focus specifically on the development of causal perception in the context of launching-event sequences because most, if not all, of the research on infants’ causal perception abilities has been conducted using standard Michottian-like launching-event sequences. I also focus on theories that were developed to provide an account of infants' developing causal perception abilities in the context of launching events because these are among the simplest events in which to observe cause-and-effect relations.

Thus, in studies that use launching-event sequences, infants typically are habituated either to a direct-launching event, a delayed-launching event, a no-collision event (sometimes called the "launching-without-collision" event), or a delayed-plus-no-collision event. In the direct-launching event, one object travels across a stage, makes contact with a stationary second object (typically located mid-screen), at which point the first object stops moving mid-screen and the second objects begins to move in the same direction as the first object. The delayed-launching event is similar to the direct-launching event except the second objects moves only after a brief delay upon contact from the first object. Likewise, the no-collision event is similar to the direct-launching event except that the second object begins to move in the absence of contact from the first object. Finally, in the delayed-plus-no-collision event, the second object begins to move both after a delay and in the absence of contact from the first object. Note that despite the fact that researchers have tended to use launching event sequences that are composed of two objects, infants' causal perception abilities have also been evaluated in the context of causal chains that are composed of three objects (e.g., Cohen, Rundell, Spellman, & Cashon, 1999). These sequences are similar to causal chains except that causal chains are composed of three or more objects whereas simple launching events are composed of two objects.

## Domain-specific theories of causal perception

The study of causal perception using launching-event sequences, particularly in adults, can be traced back to Michotte (1946, 1963). The crux of Michotte’s argument was that causes exist in the world—as reified entities—and that the human mind is wired directly to perceive them. In the context of the direct launching event in which a first object causes a second object to move immediately and upon contact, for example, Michotte argued that humans have an unavoidable tendency to see causality in such displays and to interpret the display as containing a causal agent that causes a recipient to move either immediately through direct contact, after a delay (delayed launching), at a distance (launching-without-collision), or after a delay and at a distance (delayed-plus-no-collision). On Michotte's account, the human tendency to perceive causality in launching displays such as the direct-launching event depends only the spatiotemporal relations between the objects in this sequence rather than on the particular perceptual features of each object such as its size and shape (e.g., White & Milne, 2003) and occurs automatically even for relatively impoverished stimuli (such as the launching events shown to infants that are discussed throughout this section). Michotte further argued and ostensibly showed through over 100 experiments that the perception of causality in these and similar displays is not only automatic but is irresistible and immune to top-down beliefs or intuitions. This is based on the finding that adult subjects reported rapidly seeing causality in such displays (within milliseconds) despite the fact that they were fully aware that the stimuli were impoverished and sometimes constituted just marks on a page (Michotte, 1946, 1963).

Despite never having explored causal perception in infants or young children, Michotte's emphasis on causality as being automatic, irresistible, and belief-independent is nonetheless consistent with nativist accounts that similarly argue that causal perception is present from birth or shortly thereafter (e.g., Leslie & Keeble, 1987), is automatic (e.g., Scholl & Tremoulet, 2000; Leslie, 1995), and was selected for through evolution because of the adaptive, belief-independent function it serves learners (Gelman, 1990). Furthermore, although Michotte never used the term "module" in the context of causal perception, that he placed considerable emphasis on causal perception being automatic and irresistible makes his view consistent with the modular account of cognition that was originally forwarded by Fodor (1985). However, one must interpret Michotte’s interpretation of his results with caution because his experiments were, in many cases, ill-described and his results were based on the subjective reports of his participants (e.g., Joynson, 1971). Michotte's pioneering work on causal perception nonetheless was instrumental and led to renewed attempts to study the perception of causality in infants, children, and adults (for a review see Scholl & Tremoulet, 2000).

Similar to Michotte (1946, 1963), Leslie (1995; Leslie & Keeble, 1987) has argued that infants are born with specialized mechanisms that support learning about agents and causality and that become “triggered” later in development by the appropriate input. Leslie (1995) proposed that infants are born with three innately derived modules that enable them to distinguish agents from objects on the basis of their psychological properties (theory of mind), cognitive properties, and physical properties (theory of body). This latter module is particularly important in the context of a standard launching-event sequence because it allows infants ostensibly to perceive launching events causally; that is, in terms of the transmission of FORCE—when certain conditions are met. For example, a standard launching-event sequence will be interpreted causally by a theory of body module if a first makes physical contact with a second object and subsequently transfers its FORCE to the second object such that the second object begins to move immediately and upon contact from the first object. Indeed, like Michotte, Leslie argued that the perception of causality occurs in a fixed and automatic way and is impervious to the influence of reasoning or one's belief about the events being shown to them.

Although the theory of body module is argued to be innate, it is not triggered by the appropriate input until around 6 months of age. Support for this conclusion derives from a number of Leslie's own studies in which he demonstrated not only that infants at this age can distinguish each of the launching sequences from each other (Leslie, 1984) on the basis of spatiotemporal differences, but that they possess the ability to perceive causality in these sequences. For example, 6½-month-old infants looked longer at the reversal of the direct-launching event but not at the reversal of a delayed-launching sequence following habituation to either sequence. This is because the reversal of the former, but not the latter sequence, entailed a reversal in causal roles between the first and second object (Leslie & Keeble, 1987; see Newman, Wynn, Choi, & Scholl, 2008 for a related conclusion in the context of ambiguous launching sequences).

The evidence from Leslie (1984; Leslie & Keeble, 1987) and Newman *et al*. (2008) notwithstanding, it remains unclear how the proposed modules are “triggered” by the relevant input given that information in the perceptual input often contains physical and psychological information or event what the nature of that input is (*e.g.*, Rakison & Poulin-Dubois, 2001). It is also unclear how, on Leslie’s account, a theory of body module would interface with the two remaining modules; that is, if one believes—as Fodor (1985) does—that the internal processes of each modules are impervious to outside influences, then it is unclear how information is transmitted between distinct processing modules. Note that while Leslie does not discuss how his causality modules might communicate with, and transmit information between, each other, there are a number of other perspectives that discuss how modules might pass information to and from each other (*e.g.*, Clifton & Staub, 2011; Norris, McQueen, & Cutler, 2000; Rice, Warren, & Betz, 2005) Moreover, given that most of the studies on causal perception have tended to test infants 6 months of age and older and who have had six (or more) months to learn about causality in the world, it is remains an open question whether causal perception is present at birth or emerges shortly thereafter.

Jean Mandler (1992) has argued in a similar vein to Leslie (1994, 1995) and Michotte (1946, 1963) that infants are born with a process that enables them to learn about agents, animates, inanimates, and causality. According to Mandler (1992), infants are born with a domain-general ability called *perceptual analysis* through which perceptual information is recoded into an abstract form, called image schemas, that is more accessible to infants; that is, *perceptual analysis* is what transforms perceptual information into conceptual information in which the spatial information of entities and objects is represented. *Perceptual analysis* is thought to differ from simple perception in that the former, in contrast to the latter, is an active or purposeful process of redescription whereby information in the perceptual array is attentively analyzed and abstracted to form image schemas. In terms of launching event sequences, Mandler argued that infants distinguish between causal and non-causal events by forming an image schema of “caused-to-move-inanimate” motion based on a perceptual analysis of the event sequence. Thus, infants who are shown a launching sequence may initially—through *perceptual analysis*—actively encode that the event involves two objects that move along straight paths. Infants may further encode that the second object begins to move in the presence, but not in the absence, of contact from the first object. This analysis, in turn, produces an image schema of causality in which the first object is seen as the cause of the second object’s movement. On this account, one reason infants look longer at the reversal of the causal event sequence (*i.e.*, the direct-launching event) than at the reversal of the non-causal sequence (*i.e.*, the delayed-launching event) in Leslie and Keeble (1987) was because the reversal of the first, but not the second, event violated the “caused-to-move-inanimate” motion image schema.

This ability to go beyond purely perceptual information and represent causality is important because by representing only the most important aspects of the event, infants process less information and thus have less of a biological burden placed on their information-processing abilities. Nonetheless, Mandler's theory is limited in a number of ways. First, it does not address how congenitally blind patients—who presumably are unable to see or perceive the world around them and thus who cannot engage in *perceptual analysis*—develop concepts for agents, inanimates, animates, and causality (*e.g.*, Chowdhry, 2011). Second, there is no direct empirical evidence that infants attentively analyze information in the perceptual array in the way that Mandler describes rather than by simply attending to salient properties of entities and objects (*e.g.*, motion or relative size) (*e.g.*, Slater, 1989). Third, it does not address how many times an event must be analyzed for it to be abstracted into its more conceptual, image-schematic form. Lastly, Mandler never conducted research to examine infants’ interpretation of motions other than those of drinking or giving rides (*e.g.*, Mandler & McDonough, 1996), which means that it is an open question whether *perceptual analysis* produces image schemas in putatively causal domains such as in launching-event contexts. Despite the fact that perceptual analysis represents a domain-general process, we nonetheless situate her theory within the domain-specific section because she argued that *perceptual analysis* gives rise to a domain-specific representation of causality in the form of image schemas (*e.g.*, the caused-to-move or caused motion image schema).

Still another influential domain-specific perspective is Spelke’s core systems theory (Spelke & Kinzler, 2007, 2009). The crux of her argument is that from birth infants possess a small number of separable core systems that constrain how they represent and interpret physical events. These include core systems for representing number, geometry, agency, and inanimacy. However, I will focus here on the latter two systems given their relevance to the reported projects. The core system for representing the motion of inanimate objects specifies the constraints that govern inanimate-object motion. These constraints are cohesion, continuity, and contact. The principle of cohesion states that inanimate objects maintain their boundedness and connectedness when engaged in motion; that is, inanimate objects neither spontaneously break apart when moving nor do they fuse with each other (Kestenbaum, Termine, & Spelke, 1987; Needham, 1999). The principle of continuity states that inanimate objects move continuously through time and space; that is, inanimate objects cannot spontaneously appear or disappear when moving through space, and neither can they occupy the space of another object (Kuhlmeier, Wynn, & Bloom, 2004). Lastly and perhaps most importantly from the perspective of the present experiments, the principle of contact states that inanimate objects move only when contacted by an external agent or object. Thus, inanimate objects do not engage in self-propelled motion (Spelke, 1990). It is worth mentioning that, unlike object motion, agent motion is not similarly constrained by the principle of contact. This means that agents can be self-propelled and that contact is not required for them to move from point A to point B; rather, agents represent efficient, goal-directed entities who can engage in contingent and reciprocal action and motion with other entities (Gergely & Csibra, 2003; Meltzoff & Moore, 1977; Woodward, 1998).

Support for these domain-specific positions derives from work that demonstrates: that knowledge about animates and inanimates—which have different causal affordances—can be selectively impaired following brain damage (Caramazza & Shelton, 1998; Warrington, 1981), that causal events are detected faster than non-causal events (*e.g.*, Moors, Wagemans, & de-Wit, 2017), that the preference for causal events over non-causal events may be present at birth and thus may be subserved by innate domain-specific modules or mechanisms (*e.g.*, Mascalzoni, Regolin, Vallortigara, & Simion, 2013), that by 7 months of age infants understand that objects, but not people, require physical contact to move (*e.g.*, Spelke, Phillips, & Woodward, 1995), and that specific brain regions may be involved in processing causal events (*e.g.*, Fonlupt, 2003). Despite ostensible widespread support for domain-specific mechanisms, modules, and core systems, domain-specific theories are limited for at least two reasons. First, theories that posit the existence of innately specialized modules or mechanisms suffer from a lack of direct empirical support (Rakison, 2007). Second, and perhaps most importantly, domain-specific theories are limited because they do not discuss why causal perception undergoes a developmental transition between 4½ and 10 months of age. Indeed, if it assumed that infants possess specialized mechanisms or modules that support learning about causality, then it is difficult to generate testable predictions about how, and in what ways, causal learning changes across development and to account for the existing body of research that demonstrates that causal learning does in fact change in systematic and important ways across infancy (*e.g.*, Cohen & Amsel, 1998; Oakes & Cohen, 1990; Rakison, 2005, 2006).

## Domain-general theories of causal perception

One of the first individuals explicitly to wrestle seriously with the topic of causality was Hume (1748, 1777, 1993). In particular, Hume was one of the first theorists to address what is meant by causality, whether it can be ascertained from the perceptual array, and from where it originates. In contrast to the domain-specific theorists described in the previous section, Hume adopted a domain-general perspective and was particularly interested in whether human knowledge of causal events is the product of direct experience or is present at birth. For example, Hume was interested in whether our knowledge that A causes B to move is derived online based on A’s association with B over time or whether this knowledge is known a priori. The answer that Hume offered was that all causal knowledge derives from our immediate "impressions" or direct experience with correlated causal events. Humans are able to reason, for example, that A causes B to move (as in a launching-event sequence) based on a necessary and constant connection between the two and not because of innate or a priori knowledge about A; that is, we infer that A is the cause of B's movement based on our previous experience in which we observe B constantly move upon contact from A. The reason that Hume argued that our knowledge of causality cannot be known a priori is because in the absence of any direct experience with a correlated event, there is nothing inherent about one event, object, or entity that makes it a cause or predictor of another event, object, or entity. In other words, if one has never before seen A cause B to move, there is nothing inherent to A that would enable one to predict that B will move if contacted by A. In short, Hume believed that causes do not exist in the world as reified entities but rather are inferred based on direct experience with correlated events[[1]](#footnote-1).

One of the first researchers specifically to focus on the development of causality during infancy and early childhood was Jean Piaget. In Piaget’s (1952) view, infants primarily learn about cause-and-effect in the world during the first two years of life by constructing and executing actions and observing the consequences of those actions on the self and on the world. For example, an infant who is in the sensorimotor stage may engage in a certain action—such as sucking their thumb or kicking their leg to activate a mobile that is attached to their ankle—based on the interesting and pleasurable effects that those actions confer on the self and the world. Importantly, these early representations of the causal world are limited because they are largely based on the infant’s direct interactions with the world. Piaget argued that infants move past these rudimentary representations of causality and begin to develop abstract representations of the causal world—that are not bound to self-action—only once they can reproduce the actions of others; that is, Piaget believed that a primary mechanism by which infants develop abstract representations or "images" of the causal world is imitation, an ability that is thought to change over developmental time.

According to Piaget, as infants move from one sub-stage to the next, they acquire the ability to imitate increasingly complex actions. This ability to imitate increasingly complex actions overtime becomes internalized and leads to the formation of “images” or concepts of the causal world, which are thought to emerge around the age of 6 or 7. Piaget’s perspective can be considered a domain-general theory because such mechanisms (*i.e.*, imitation, assimilation, accommodation) are thought to support learning in a range of content areas, including in causal contexts. Nonetheless, Piaget’s theory of causality suffers from a number of limitations, which has led to considerable criticism (*e.g.*, Sugarman, 1987). These criticisms range from inadequate study methodologies, to how he underestimated children’s learning abilities (*e.g.*, Baillargeon, 1987; Spelke & Newport, 1998).

Proponents of the information-processing perspective have also forwarded a similar argument to that of Piaget. The crux of their argument is that causal perception is underpinned by an associative, information-processing learning mechanism rather than by innate mechanisms, modules, skeletal systems, or core systems (e.g., Oakes & Cohen, 1990; Oakes & Cohen, 1994; Cohen et al., 1998; Rakison, 2005; Rakison & Lupyan, 2008; Rakison, Smith, & Ali, 2016). On this account, the ability to perceive causality in launching events develops incrementally and results from changes in information-processing mechanisms and processes that are present from birth and that support learning in a range of content areas. On this view, infants learn to perceive causality in launching events first by attending to the low-level features and then, only later, by combining those features with higher-level features until a complete causal concept has been formed. For example, young infants shown the launching-event sequences described earlier would have an initial bias to encode only the low-level perceptual features of the objects involved such as their color, shape, or path of movement. This assertion is based on the finding that 4-month-olds, but neither 5½-month-olds nor the 6¼-month-olds, looked longer at the causal event regardless whether they were habituated to the causal event or either of two non-causal events. Cohen and Amsel (1998) interpreted this finding to mean that the youngest infants were entrained on the continuous motion of the direct-launching sequence and recovered attention to non-contingent motion (see also Belanger & Desrochers, 2001).

However, as working memory and processing speed continues to improve, proponents of the information-processing perspective argue that infants begin to process the independent features of launching-event sequences but not necessarily the causality in these sequences. The reason that infants at this stage are not expected to process launching events in terms of causality is because such perception requires that infants be able to encode and integrate low-level perceptual features (*e.g.*, continuous motion) with higher-level independent features (*i.e.*, spatial and temporal relations between a first and second object; for a computational instantiation of this account see Cohen, Chaput, & Cashon, 2002). This implies that infants between 4½ and 6½ months of age should look as long at the reversal of a direct-launching sequence as they should at the reversal of a delayed-launching sequence because both reversals entail identical spatiotemporal changes. Support for this conclusion was also garnered from Cohen and Amsel (1998) who showed that 5½-month-olds habituated to one of the two non-causal events (*e.g.*, no-collision) showed elevated looking to the other non-causal event (*e.g.*, delayed launching) and the causal event (*e.g.*, direct-launching) than to the familiar non-causal event (*e.g.*, no-collision); that is, these infants looked longer at the test sequences that violated the habituation sequence in terms of the spatial and/or temporal cues (*i.e.*, the independent features).

The ability to combine independent features to process causality in launching events and distinguish causal from non-causal events would be expected to emerge still later on this account. This is because infants would have had more time to learn about causality and their information-processing abilities would have time further to improving. For example, Cohen and Amsel (1998) showed that 6¼-month-olds habituated to either of the non-causal events dishabituated to the causal event, whereas those habituated to the causal event dishabituated to both non-causal events. Evidence that infants 6¼ months of age and older respond to the causality of launching event sequences—after having presumably combined the independent features—also derives from research by Leslie and Keeble (1987) who showed that 7½-month-olds looked longer at the reversal of a causal sequence than at the reversal of a non-causal sequence and that 10-month-olds habituated to the causal event—which used realistic stimuli—will dishabituate to non-causal events, whereas 10-month-olds habituated to either of two non-causal events will dishabituate to the sole causal event (*e.g.*, Oakes & Cohen, 1990; for evidence that infants can perceive causality earlier than 6 months of age when given causal action experience see Rakison & Krogh, 2012).

The theories and research outlined in this section differ from each other in important ways but all have in common the notion that learning about the world is underpinned by mechanisms that support learning in a range of content areas (*e.g.*, associative learning, habituation, conditioning, and imitation). For Hume, learning about causality involved observing "constant conjunctions" or correlated events. For Piaget, learning about and developing abstract representations of the causal world involved imitating the actions of others and objects. For Cohen (Cohen *et al.*, 1998), learning is thought to involve an ever-improving information-processing system that supports learning about causality, angular relations, object unity, and complex patterns (*e.g.*, Cohen & Amsel, 1998; Cohen & Younger, 1984; Cohen, Chaput, & Cashon, 2002; Younger & Cohen, 1983, 1986; Rakison, 2005, 2006).

Although the domain-general position has been criticized for being underconstrained (*e.g.*, Keil, 1981) and at odds with research that suggests that the brain is composed of specialized processing areas (*e.g.*, Baron-Cohen *et al.*, 1994), this position is supported by research that demonstrates that knowledge representations may be instantiated in distributed networks in the brain (Humphreys & Forde, 2001; Rogers & McClelland, 2004); that damage that was thought originally to be restricted to one area may, in fact, be more widespread and may span multiple domains especially in children in which specialization and localization of particular functions have not been established (*e.g.*, Karmiloff-Smith, 2009; Karmiloff-Smith, Scerif, & Ansari, 2003); that early learning constrains later learning (*e.g.*, Rakison & Poulin-Dubois, 2002); and that these constraints need not be innately specified but can emerge based on experience in the world, based on maturing brain systems, and based on the same underlying mechanism that processes input from multiple domains and at increasing levels of sophistication (*e.g.*, Madole & Oakes, 1999; Rakison & Lupyan, 2008; Rakison & Yermolayeva, 2011).

## Summary

There has been considerable disagreement among researchers about the nature of the representations that underlie causal perception, the mechanisms that support this ability, and about the extent to which this ability changes over developmental time. On the one hand—in the spirit of Michotte—it has been proposed that infants acquire knowledge about causality via domain-specific modules or mechanisms that are thought to be present from birth or shortly thereafter. On the other hand—in the spirit of Hume—others have argued that causal perception is an ability that develops incrementally and is subserved by a domain-general learning mechanisms that support learning in a range of content areas.

The goal for the following section will be to discuss the different theoretical views of causal reasoning and the research that has been taken to support these perspectives. The theoretical views that we will discuss in particular are associative views of causal reasoning, contingency-based views of casual reasoning, and Bayesian inference as causal reasoning view that encompasses the causal graphical model approach. Special attention will be given to causal graphical models because it is currently the predominant theory of causal learning and is the principal theory of which the present studies were designed to test the viability.

# Studies on causal reasoning and corresponding theories

## Associative models of causal reasoning: The Rescorla-Wagner model

One of the earliest models to be proposed to explain associative causal learning was the Rescorla-Wagner model (henceforth, the R-W model; Rescorla & Wagner, 1972). Although this model was developed originally to account for classical condition in non-human animals, it has since been extended to account for causal learning in humans. According to the model, the extent to which a reasoner learns the associative relationship between a conditioned stimulus and an unconditioned stimulus in a Pavlovian learning context or between a cause and an effect in a causal context depends on how well the conditioned stimulus or cause predicts the conditioned response or effect. For instance, if during the course of learning a given cue (e.g., light) does not predict an outcome (e.g., shock) to the level that it should such that there is a discrepancy between what is expected and what actually occurs, the "weight" or associative strength between that cue and the outcome will be modified until the cue predicts the outcome at or around asymptote, although this tends not to be true in most experimental designs (e.g., Danks, 2003).[[2]](#footnote-2)

Despite the fact that the Rescorla-Wagner model can account for a number of well-known phenomena such as blocking (e.g., Kamin, 1969) and conditioned inhibition (e.g., Wagner et al., 1980; see Chapman, 1991 for retrospective reevaluation variants), it nonetheless suffers from several limitations (see Miller, Barnet, & Grahame, 1995 for an extensive review of these limitations). First, it does not discriminate between cues that are merely correlated with an effect and cues that produce effects. This is because there is nothing inherent in the model that enables it to distinguish between an actual cause and an unobserved common cause given that the weights that underpin the association between a cue and an outcome merely represent the strength of the associative relation. Second, the RW model does not allow for single-trial learning, although it is possible to learn to predict an outcome at asymptote—after two trials of learning—when the learning rates in the model are each set to their maximal values. This is problematic for the model because recent evidence suggests that children and adults can make causal inferences and design novel interventions to change the likelihood of certain relationships from a single learning trial or a small number of trials (e.g., Sobel et al., 2004; Griffiths et al., 2011). Third, the RW model is unable to explain why adults make different causal judgements based on whether cues represent causes or effects. For example, Waldmann (2000) showed that participants treat competing cues differently depending on whether they are described as causes of an effect (common-effect model) or effects of a cause (common-cause model): when competing cues were described as causes, participants were more likely to block the redundant cue. However, when competing cues were described as effects, participants were less likely to block redundant effects (see also Kloos & Sloutsky, 2013). The RW model is unable to account for this finding because it predicts blocking regardless whether the input given to the model represents causes or effects.

Finally, it is limited because it cannot account for why non-human animals, children, and adults engage in backwards-blocking reasoning in causal contexts (Blaser, Couvillon, & Bitterman, 2004; Griffiths et al., 2011; Shanks, 1985; Shanks & Dickinson, 1988; Sobel et al., 2004). In a typical backward-blocking (hereafter, BB) trial, participants are shown events that follow an AB+ A+ format and are then asked to provide a causal rating for both objects; that is, objects A and B are both shown to produce an effect, after which object A by itself produces the effect. These ratings are then compared to the ratings of A and B provided by participants who see an AB+ A- sequence (called indirect screening-off; henceforth IS). The results of these studies typically show that children and adults rate B less causally in the BB condition than B in the IS condition (or to a control condition in which participants are shown AB+ only). The reason subjects are thought to rate B less causally in BB trials than IS trials is because A is shown explicitly to produce the effect. This basic finding has been replicated a number of times (e.g., Carroll, Cheng, & Lu, 2013; Griffiths, Sobel, & Tenenbaum, 2011) and is often interpreted to highlight the failure of rudimentary associative models such as the RW model. This is because the RW model predicts the association between B and the outcome to be the same across both conditions. In other words, in the BB condition, as in the IS condition, B is never shown by itself and only occurs once, which leads to the prediction that the rating of B should be the same in both the BB and IS conditions. It should be noted that although the RW model is criticized for being unable to account for backwards-blocking reasoning in humans and some non-human animals, the evidence is in fact mixed about whether and to what extent humans engage in backwards-blocking reasoning. For example, using a two-stage blocking procedure, Larkin, Aitken, and Dickinson (1998) found that subjects did not block the redundant cue following an AB+ A+ sequence, and Lovibond et al. (2003) extended this finding to demonstrate that whether human reasoners block redundant cues depends on the extent to which they believe that causes combine additively to produce larger effects.

## Associative models of causal reasoning: The modified Rescorla-Wagner model

Based on the limitations of the RW model, researchers have proposed modified associative models that have been able to account for the BB finding, although the details of each model differs (*e.g.*, Blaisdell, Bristol, Gunther, & Miller, 1998; Dickinson & Burke, 1996; Dwyer, Mackintosh, & Boakes, 1998; Larkin *et al.*, 1998; see also De Houwer & Beckers, 2002 for a more general review of these associative models). One such successful model is Van Hamme and Wasserman's (1994) modified RW model. Although the RW model and its modified variant use learning-rate parameters to adjust associative strengths between cues and outcomes, these models differ in how these parameters are interpreted for present and absent cues. Note that in both models, α corresponds to the salience of the cue or the candidate cause whereas β corresponds to the salience of the outcome or effect. In the traditional RW model, β is largest when the outcome occurs than when it does not occur, whereas α is 0 when the cue is absent but is set to a non-negative value between 0 and 1 when the cue is present. The reason α is set to 0 when cues are absent is because it is assumed that learning cannot occur for absent cues. In this way, absent cues are assumed to espouse no salience. This decision to set α to 0 is nonetheless problematic in the context of the backwards-blocking design because the traditional RW model erroneously predicts that participants' ratings of B after the AB+ trial should not differ from their ratings it after the A+ trial. This is at odds with the findings that adults rate B lower following the A+ trial than after the AB+ trial and compared to a control condition in which adults are only shown the AB+ compound (*e.g.*, Griffiths *et al.*, 2011; Lovibond *et al*., 2003; cf., Larkin *et al.*, 1998), they will block B when they are told that A and B represent causes but not when both cues represent effects (Kloos & Sloutsky, 2013; Waldmann, 2000), and the finding that 4-year-old children classify B as a cause following an AB+; A- event than after an AB+ A+ event (e.g., Sobel *et al.,* 2004).

Van Hamme and Wasserman (1994) addressed this limitation by assuming that cues that are unexpectedly absent take on negative salience rather than no salience. This key assumption enabled the modified RW model to account for the backwards-blocking finding as well as related findings for which the traditional RW model fails to account and has even been instantiated in the context of connectionist computational models to account for the inverse base-rate phenomenon in a category-learning context (e.g., Chapman, 1991; Shanks, 1985). Despite the fact that the modified RW model is able to account for the backwards-blocking finding as well as other related findings, the model suffers from many of the same shortcomings as the traditional RW model. For example, similar to the RW model, the modified RW model is unable to distinguish between cues that are merely correlated with effects from cues that produce effects. Moreover, the modified RW model is limited because it requires many learning trials takes for reliable associations to be established between cues and outcomes. Finally, it is limited because it cannot account for why adults treat cues differently depending on whether they represent causes or effects. These limitations as well as others have led many researchers to reconsider the viability of these models as models of human causal learning.

## Contrast-contingency models of causal reasoning

Given the limitations of the associative models presented in the previous section, some researchers have proposed alternative models—called contrast or contingency models—that are purported to provide a better account of human causal learning. One such model is Jenkins and Ward's (1965) traditional contingency model in which human reasoners are argued to compute contrast statistics that enable normative causal inference. In particular, Jenkins and Ward (1965) argued that reasoners use the difference between *P*(*E*|*C*) and *P*( ¬*E*|*C*) to infer causal relationships. Here, the first and second terms correspond respectively to the probability of outcome E given a candidate cause C and the probability that an outcome will not occur given the same candidate cause. It is generally assumed that if the difference between these two terms is positive—that is, if *P*(*E*|*C*) - *P*( ¬*E*|*C*)) > 0—then the candidate cause is assumed to be generative. In contrast, if the difference between the two terms is negative, then the candidate cause is assumed to be preventative. Finally, if the difference between the two terms if zero—that is, the effect occurs equally often in the presence or absence of the cue—causality is not attributed to the candidate cause.

Support for this model derives from studies that demonstrate that human reasoners provide causal ratings that generally are consistent with the predictions of the model. Wasserman et al. (1993) demonstrated, for example, that subjects’ causal ratings correlated highly with objective contingency in a causal task that required subjects to determine to what extent tapping a key caused a distant light to illuminate. Similarly, Allan and Jenkins (1980) found that when the number of possible actions matched the number of possible outcomes, subjects provided accurate causal ratings that were consistent with objective contingency: compared to subjects who were asked to determine whether engaging two possible actions (i.e., moving a joystick in one of two directions) produced a single effect (i.e., caused a picture of the Loch Ness Monster to appear), subjects who engaged in a single action to produce the single effect provided causal ratings that were consistent with objective contingency. This finding was replicated by Allan and Jenkins (1983) in a novel task in which subjects were asked to judge whether moving a joystick in one or more direction caused a centrally located dot to move in one of two directions: subjects who were asked either to determine to what extent one of two possible actions (i.e., moving a joy stick in one of two directions) produced one of two possible outcomes (i.e., either a dot moved from the center of a screen upward or from the center of a screen downward) or to determine to what extent a single action produced a single effect provided causal ratings that accorded with objective contingency. The authors interpreted this finding to mean that humans reasoners will track accurately objective contingency when the relation between causes (or actions) and effects is compatible; that is, when the number of possible actions matches the number of possible effects. Indeed, Allan and Jenkins (1983) found that when the number of possible actions did not match the number of possible effects, adults were much less accurate at tracking objective contingency. Finally, Alloy and Abramson (1979) showed that other factors also contribute to the accuracy with which human reasoners track objective contingency. In particular, they demonstrated that depressed subjects were more accurate at tracking objective contingency than non-depressed subjects in a task in which subjects were asked to determine whether pressing a button caused a green light to illuminate. Alloy and Abramson argued that this finding emerged because depressed subjects compared to non-depressed subjects are less likely to overestimate—and thus more likely to provide accurate causal estimates—the degree of contingency between their own actions and effects perhaps because of learned helplessness. These findings illustrate that human reasoners are capable of tracking objective contingency, although it depends on important contextual and dispositional factors.

Despite these findings, the model suffers from several notable limitations. First, the traditional contingency model provides no way to distinguish between correlational and causal relations. For example, although sneezing might follow regularly from putting one's hand to one's mouth, it would be incorrect to conclude that the act of putting one’s hand to one’s mouth causes sneezing despite the apparent contingency between the two. Second, it fails to account for a number of subtle but important findings in the literature. For example, it fails to explain why a person's dispositional state affects their judgements of and sensitivity to objective contingency between actions and outcomes (e.g., Alloy and Abramson, 1979); why subjects are better at tracking objective contingency when the relationship between actions and outcomes is compatible than when it is incompatible (e.g., Allan & Jenkins, 1983; Jenkins & Ward, 1965); why judgements of objective contingency correlate highly with how frequently the outcome occurs regardless of the amount of actual control reasoner exerts over outcomes (e.g., Allan & Jenkins, 1983); why subjects judged negative correlations as being closer to zero than positive correlations even when objective contingency was held constant (e.g., Erlick & Mills, 1967); why the causal relation between actions and outcomes is judged to be stronger (i.e., the action is said to be more causal) when that relation is positive than when it is negative (e.g., Allan & Jenkins, 1983); and why judgements of causal contingency are more accurate when the relationship between actions and outcomes is provided in the form of a summary table rather than when it is shown directly or is presented as an event timeline (e.g., Smedslund, 1963; Jenkins & Ward, 1965; Wasserman & Shaklee, 1983). Given these limitations, the traditional contingency model must be interpreted cautiously as a viable model of human causal reasoning.

## The Power PC Theory

Based on the limitations of the traditional contingency model, Cheng (1997) proposed the Power PC theory of causal reasoning. The Power PC model is similar to the traditional contingency model in that both models assume that learners infer causal relations on the basis of contingency information. However, the Power PC model differs from the traditional contingency model in that it uses two estimates—which are Δ*P* and (1-*P*(*E*|¬*C*)) and which are expressed as a ratio—not only to determine whether a causal relation exists between candidate causes and effects but to determine the causal strength of that relation. Thus, according to the Power PC model, the causal power of a generative candidate cause can be directly estimated as:

*p*c =

where the denominator provides an estimate of the probability of the effect given the cause (taking into account all alternative causes). Note that the equation for preventative candidate causes is similar except that numerator is expressed as -Δ*P*. It should also be clear from the above equation that Δ*P* underestimates causal power *pc*. This is because Δ*P*, but not *pc*, does not account for alternative causes and because it does not take into account the base rate of the effect independent of the likelihood of the effect given candidate causes. Thus, if *P*(*E*|*C*) = 0.95 and *P*(*E*|¬*C*) = 0.25, then Δ*P* = 0.7 and *pc* = 0.93.

It is worth noting that Δ*P* is an exact estimate of causal power if, and only if, *P*(*E*|¬*C*) = 0. However, the model predicts that learners should provide increasingly conservative estimates of C as *P*(*E*|¬*C*) approaches 1, but should be uncertain about the causal power of a candidate cause when *P*(*E*|¬*C*) = 1. This model extends, and hence improves upon, the traditional contingency model because it can account for the base rate effects described earlier for which the traditional contingency model is unable to account; that is, it can account for why judgements of objective contingency sometimes increase as the frequency of the outcome increases. For example, given a contrast of .1 (i.e., *P*(*E*|*C*) - *P*(*E*|¬*C*) = .1), participants will provide higher causal ratings if the base rate of the effect in one situation (e.g., *P*(*E*|¬*C*) = .65) is higher than that in another situation (e.g., *P*(*E*|¬*C*) = .25). Despite being able to account for the base rate of the effect and a number of other important findings (for a review, see Cheng, 1997), the model suffers from several limitations. First, the Power PC model—as well as both the traditional contingency model and the Rescorla-Wagner model—makes no predictions about how causal judgements should change from one learning trial to the next. This is because contrasts are computed over a series of causal events rather than after individual events; that is, the model assumes that causal decisions are made at asymptote. Second, although the traditional contingency and Power PC models account for the base rate of the effect, neither the traditional contingency model nor the Power PC theory accounts for the base rate of candidate causes themselves. This is a key limitation because recent research demonstrates that children (e.g., Sobel et al., 2004) and adults (e.g., Griffiths et al., 2011) are sensitive to the base rate of the causal candidate itself (in addition to the base rate of the effect in the absence of candidate causes). Third, the Power PC model, similar to both the RW and the traditional contingency models, is unable to explain how it can be that children and adults make causal decisions based on a limited number of learning trials; that is, because the Power PC model, the traditional contingency model, and the RW model all assume that causal decisions are made only once the set of all causal events have been experienced, all three models are unable to account for the "one-shot" learning observed in recent causal studies (e.g., Gopnik et al., 2001; Sobel et al., 2004). Finally, and as mentioned above, all three models are limited because evidence is mixed about whether, and to what extent, human learners can accurately compute conditional contrasts in the first place (e.g., Alloy & Abramson, 1980; Jenkins & Ward, 1965; Wasserman & Shaklee, 1984). Based on these limitations, the Power PC theory, much like the traditional contingency model and the RW model, should be interpreted cautiously as a viable model of human causal learning.

## The Bayesian Inference perspective

In light of the limitations of associative and contingency-based models, in recent years some researchers have argued that human causal reasoning and learning can be best explained within the Bayesian-inference framework. The crux of this perspective is that infants, children and adults possess from birth an evolved causal system that reconstructs the causal relations in the world based on observed data in much the same way that the human visual system reconstructs the visual environment from two-dimensional retinal input (e.g., Gopnik et al., 2004; Gopnik & Wellman, 2012). In addition to possessing a causal system that is argued to be present from birth or shortly thereafter, proponents of this perspective maintain that human learners use Bayesian inference—that is, a simple form of Bayes' rule—to choose the causal hypothesis that is most consistent with the observed data (for a detailed explanation of how humans use Bayes' rule to choose the most likely hypothesis see Gopnik et al., 2004).

This perspective has garnered considerable attention in the literature for a number of reasons. First, it provides a better account of the base-rate findings discussed in the previous section for which associative- and contingency-based models fail to account. For example, this perspective can account for the finding that children (e.g., Sobel et al., 2004) and adults (e.g., Griffiths et al., 2011) attribute greater causality to causes that are common than to causes that are rare. This is because Bayesian models—but neither contingency-based models nor associative-based models—incorporate priors that encode the likelihood of a particular hypothesis independent of observed data. Second, unlike associative- and contingency-based models that require that large amounts of data be experienced before accurate causal inferences can be made, this perspective accounts for the finding that children and adults make causal inferences and design novel interventions to generate different outcomes based on a single or small number of learning trials (e.g., Sobel et al., 2004; Griffiths et al., 2011); indeed, many fewer trials than would be required if adults reasoned according to associative or contingency models. This is because hypotheses in a theoretical hypothesis space—over which Bayesian inference is thought to operate—are updated on a trial-by-trial basis rather than after a large number of learning trials have been presented. Third, this perspective explains how causal inferences are made from purely associative and correlational data based on causal interventions that can discriminate between competing causal hypotheses. In this way, the Bayesian inference perspective but neither the associative nor contingency perspectives is able to account for the conditions under which learners can infer that a causal relation exist from correlational input.

Despite the ability of this perspective to account for the causal-learning abilities of infants (e.g, Sobel & Kirham, 2006, 2007; cf., Benton & Rakison, 2016), children (e.g., Gopnik et al., 2001; Sobel et al., 2004), and adults (e.g., Griffiths et al., 2011), the notion that humans use Bayes' rule to reason about causal events is problematic for a number of important reasons. The first problem is that it is unclear where priors come from and what the nature of the mechanism or set of mechanisms is that underpins them. Although some authors have argued that priors track the statistics in the environment and, as such, are ecologically valid (e.g., Griffiths & Tenenbaun, 2006), it is as yet unresolved whether such priors result from a domain-general associative learning mechanism (e.g., Rakison & Lupyan, 2008) or from domain-specific modules, mechanisms, and systems that support causal learning and inference (e.g., Gopnik et al., 2004; Leslie, 1994, 1995).

The second problem with this perspective concerns that notion of posterior probabilities. In the context of Bayes’ rule, the posterior probability is what learners are said to compute to determine the most likely hypothesis and is proportional to the product of the prior probability and the likelihood based on observed data—that is, *P*(*H*|*D*)∝*P*(*H*)*P*(*D|H*) (e.g., Gopnik et al., 2004). The notion that humans represent the entire hypothesis space, much less compare and update each hypothesis with new data, is nonetheless problematic because it is computationally intractable and biologically implausible; in fact, this problem has been called the “search problem” (e.g., Gopnik & Wellman, 2012; Gopnik et al., 2004). If there are a small number of potential causes, then the problem of enumerating over all hypotheses and comparing the data to each hypothesis is a relatively trivial one. However, the problem becomes apparent as the number of potential causes increases. Indeed, as the number of candidate causes increases, the number of potential comparisons grows superexponentially. Nonetheless, solutions to the search space problem have been offered. For example, some have argued that humans use a version of “simulated annealing”—in which one’s search strategy “cools off” or becomes less noisy with experience—to reduce the search space (e.g. Lucas, Bridgers, Griffiths, & Gopnik, 2014; cf., Frank, Goldwater, Griffiths, & Tenenbaum, 2010). Others have argued that human reasoners adopt a win-stay-lose shift strategy—in which the “winning” hypothesis is the one that is most consistent with the data—to reduce the hypothesis space (e.g., Bonawitz, Denison, Griffiths, & Gopnik, 2014; Denison, Bonawitz, Gopnik, & Griffiths, 2010). Still others have argued that merely explaining a causal system can help to reduce the hypothesis space because explanations tend to support generalizations with the broadest scope (e.g., Walker, Lombrozo, Legare, & Gopnik, 2014). However, these accounts are limited because they lack biological plausibility, make untested assumptions about the nature of the underlying causal representation, or lack direct empirical support.

The third problem is that it is an open question whether causal events are represented as causal graphical models (discussed in detail below) or learned associations between low-level perceptual features. This is an especially acute problem because if it is argued that humans use Bayes' rule to reason over graph-like causal structures, then it must be the case, and must be shown empirically, that causal events are represented as causal graphs in the first place. Given these key limitations, this account must be interpreted cautiously as a viable model of human causal learning.

## Causal Graphical Models

As was discussed in the previous section, proponents of the Bayesian inference perspective maintain that human learners use Bayes' rule to choose the causal hypothesis that is most consistent with, and thus is the most likely to have generated, the observed data. In particular, it is argued that learners represent causal events as causal-graphical models and that they use Bayes' rule to choose the CGM that has the highest posterior probability; that is, the CGM that is most consistent with the causal data.

A CGM is an abstract representation of the causal relations among and between random variables. These random variables—which are represented as nodes in a CGM structure—may represent objects, events, or entities and encode probability distributions such that a change in the probability distribution of a first variable will alter the probability distribution of a second variable assuming a dependence between both variables. These random variables are connected to other random variables via directed edges (i.e., arrows), which encode causal directionality. In Figure 1A, for example, A is connected to B by an edge that is directed into B. That the arrow is directed into B from A and not the reverse indicates (a) that A is the cause of B and (b) that A is the parent of B, which is the child of A.

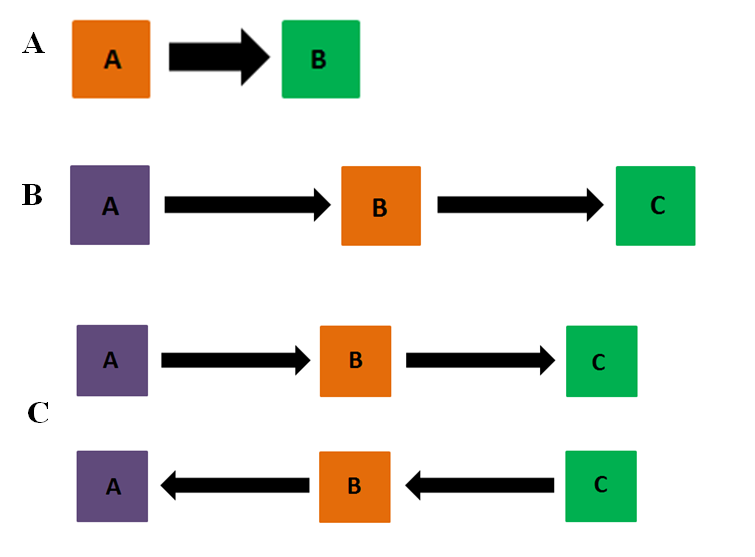


Fig. 1. Examples of (A) a standard launching event sequence that is composed of two objects, (B) a single causal chain that is composed of three objects, and (C) two causal chains each of which are composed of three objects.

If a third random variable was added to the CGM in Figure 1—such as in Figure 1B—such that A is connected to B via an edge that is directed into B from A and B is connected to C via an edge that is directed into C, then A is the ancestor of C, B is the parent of C, and C is the child of B and the descendent of A. For graphs in which, for example, A is the sole variable, A is simultaneously its own ancestor and descendent. Finally, because A is the sole parent in Figures 1A and 1B, A is said to be an exogenous variable. However, B in Figure 1A and B and C in Figure 1B are said to be endogenous variables because edges are directed into each of them and, as such, their states are set by their parents. In general, altering the probability distribution of exogenous variables will alter the probability of the endogenous variables to which it is directly or indirectly connected.

CGMs are of particular import for two reasons. First, as was mentioned in the preceding section, they support causal intervention. Indeed, some researchers have even argued that intervention is the *sine qua non* of causal reasoning because it is a viable mechanism that supports theory formation (e.g. Gopnik & Schulz, 2004), it enables learners to determine the effects of interventions on particular variables (e.g., Gopnik et al., 2001; Sobel et al., 2004), to determine whether and to what extent an intervention on one variable affects another (e.g., Steyvers et al., 2003), to determine what would have happened if a different intervention than the current one was undertaken (e.g., Sobel, 2004; Harris et al., 1996), and to distinguish correlational relationships from causal relationships which is the goal of science and for learning how the world works. Thus, if one believes, but cannot be certain, that the CGM in Figure 1A describes the true causal relation between the two variables, he or she can choose to intervene on A to determine whether such an intervention changes the state of B (or vice versa).

Second, CGMs are important to consider here because one of the goals of the present series of experiments is to determine whether and to what extent humans represent causal events as CGMs or learned associations between low-level perceptual cues. I have written elsewhere (e.g., Benton & Rakison, under review) that causal perception and causal reasoning are underpinned by the same underlying domain-general associative learning mechanism and that both abilities rely on, and make use of, the same sources of perceptual cues and features. The crux of this account is that infants, children, and adults—via a domain-general associative learning mechanism—represent causal events in terms of: (a) the cues of space, time, and frequency (i.e., how often the effect happens in the presence of a candidate cause) as well as in terms of the perceptual—or surface—features of objects that participate in causal events and (b) the associative relation among and between these cues. Thus, rather than causal perception and causal reasoning being functions of a causal schema (e.g., Mandler, 1992), innate causal modules (e.g., Leslie, 1994, 1995; Newman, Choi, Wynn, & Scholl, 2008), skeletal causal systems (e.g., Gelman, 1995), or CGMs (e.g., Gopnik et al., 2004), humans’ representations of causal events rely on and make use of the associative relation between low-level perceptual cues and features.

In terms of a causal launching sequence that is composed of two objects, for example, the view is that learners associate the movements of the second object—including whether this objects moves after a delay or in the presence of a spatial gap—with the cessation of movement of the first object. In addition, because the surface features of the first and second objects presumably differ in such sequences—though not always (see Newman et al., 2008 and Scholl & Nakayama, 2004 for examples of exceptions)—the perceptual identity of the second object is associated with the movement of the second object and likewise for the first object. In principle, this association could then lead to expectations on the part of the infant, child, or adult about the identity and kind of motion that a second object will engage in when the first object is presented and about the causal identity of both objects—that is, whether they are agents or recipients. This associative link is important from the perspective of prediction because the presence of one object—which looks and moves in a certain way—will lead to an expectation about whether or not the second object will move in the presence of a delay or gap, about the order in which both objects will begin to move, about the identity of the second object, and about whether or not the event is causal based on the presence or absence of contiguity relations (see Rakison & Lupyan, 2008 for a closely related idea) and can inform later causal interventions (e.g., Sobel et al., 2004). In plain, rather than causal events being represented as CGMs, I argue that causal events are represented as learned associations—which encode, spatial, temporal, covariational, and perceptual cues—between objects and entities.

Second, and perhaps most importantly in the context of the present experiments, CGMs are important because they instantiate conditional independence relations between variables such that some variables will be probabilistically independent of other variables conditioned on the state of their parents. In Figure 1B, for example, C is probabilistically independent of A conditioned on its parent B. This aspect of CGMs is known as the causal Markov condition and can be expressed formally by the following equation that defines a joint-probability distribution over all *n* state variables:

= ,

where each term, *P*(*Xi*|parents(*Xi*)), encodes local causal relations between one node *Xi* and its parents, parents(*Xi*), and where the value of node *Xi* is a function of the states of its parents(*Xi*). Note that because each node in a CGM instantiates a particular probability distribution, the equation above defines a joint probability distribution over all *n* state variables in a CGM.

The Markov condition is important for yet another reason; that is, learners who are sensitive to and can encode the Markov condition in causal sequences should be unable to distinguish sequences that instantiate the same condition based on purely observational data and assuming both that the joint-probability distributions are multivariate Gaussian and that the variables that comprise a sequence are discrete-valued. Sequences that instantiate the same condition—that is, that encapsulate the same full set of conditional independencies—such as the sequences in Figure 1C are said to form Markov-equivalence classes. Thus, if learners are sensitive to the Markov condition and can recognize when causal structures instantiate the same condition, then they should treat both sequences in Figure 1C as equivalent because they express the same set of conditional independencies. Learners should also treat both sequences equivalently because both structures produce the same patterns of data; namely, in the absence of important cues to causal structure such as temporal priority or contiguity, both sequences will produce either of two patterns of data: either all of the nodes will be on or none of the nodes will be on in both sequences and hence should be indistinguishable on the basis of pure observation.[[3]](#footnote-3)

Evidence that children are sensitive to the Markov condition ostensibly derives from research that has used the blicket-detector paradigm. Note that a blicket detector is a machine lights up and plays music when blickets are placed on its surface but not when non-blickets are placed on it. Although children’s causal reasoning abilities have been tested in different conditions using this paradigm, the two conditions I will focus briefly on here are the indirect-screening off condition and the one-cause condition. This research has shown that in the indirect-screening off condition—in which children first observe that objects A and B together make the blicket detector light up when placed on the surface of the detector and that A fails to make the detector activate when placed alone on the surface—4-year-olds will place object B, but not object A, on the detector (e.g., Sobel et al., 2004). Likewise, in a one-cause condition—in which children observe that A alone makes the detector activate, that B alone fails to make the detector activate, and then that A and B together make the detector activate—4-year-olds will place object A, but not object B, on the detector when asked to make it go (e.g., Gopnik et al., 2001). Sobel et al. (2004) and Gopnik et al. (2001) interpreted these findings to mean that children can encode conditional independencies in causal event sequences; that is, children are sensitive to the Markov condition because they recognized that the effect was independent of A conditional on B in the indirect-screening off condition and that the effect was independent of B conditional on A in the one-cause condition.

However, these results should be interpreted cautiously for four reasons. First, it is unclear whether the 4-year-olds would be sensitive to the Markov condition if the experimenters did not use causal language to describe the task. For example, the children were told that only blickets make the machine go and that their job was to determine which objects were blickets based on different patterns of activation. The use of such causal language could well have directed children's attention to the conditional-independence relations (i.e., to the Markov condition) between the candidate blickets in the indirect screening-off and one-cause conditions. Thus, it is an open question whether children can encode the Markov condition and can form Markov equivalence classes when non-causal language is used to describe otherwise causal tasks.

Second, because the primary goal of the studies by Sobel et al. (2004) and Gopnik et al. (2001) was to examine how children reason about causes in different conditions rather than to examine children’s sensitivity to the Markov condition per se, one must be cautious about concluding that children are sensitive to the Markov condition; that is, because the Markov condition was not experimentally manipulated in either study, no strong conclusions can be made about whether, and to what extent, children are sensitive to the Markov condition.

Third, it is an open question whether children's sensitivity to the Markov condition is robust to low-level perceptual changes such as changes to the spatiotemporal features of the causal events; that is, it is unclear whether children’s sensitivity to the Markov condition will persist when salient perceptual manipulations—such as manipulations to the spatiotemporal relations between and among elements—are made to the causal events about which they are asked to reason. Despite the fact that there is a paucity of research to clarify this issue, there is reason to believe that children’s representation of causal events may be grounded in low-level perceptual features rather in terms of the underlying conditional-independence relations (i.e., that Markov condition). One study by Kushnir and Gopnik (2007), for example, showed that 3-year-olds are less willing than 4-year-olds to attribute causality to objects that generate effects at a distance than to objects that generate effects through physical contact.

Finally, it is as yet unknown whether children are sensitive to Markov-equivalence classes. Indeed, although the research by Sobel et al. (2004) and Gopnik et al. (2001) demonstrated that children may be sensitive to the Markov condition, neither study tested whether and to what extent children are sensitive to Markov-equivalence classes. This is because in Sobel et al.’s (2004) and Gopnik et al.’s (2001) experiments children were asked only to reason about a single causal event rather than multiple causal events. The reason it is important to use multiple causal events is because it is possible to test to what extent any two (or more) events are Markov-equivalent.

Evidence that adults may be sensitive to the Markov condition can be found in Steyvers et al*.* (2003; see also Sobel & Kirkham, 2006, 2007 and Gopnik *et al.*,2001 for work with infants and children on this issue). They demonstrated that adults were slightly better than chance at distinguishing structures from different equivalence classes based on passive observations of the pattern of communication between three aliens but slightly better when they were allowed to intervene on these systems. In particular, following a round of trials in which subjects observed two outer aliens read the mind of a middle alien (i.e., common-cause model) or a middle alien read the minds of either of the two outer aliens (i.e., common-effect model), Steyvers *et al.* (2003) found that adults were 18% accurate at choosing the correct structure that generated the data—which was slightly better than chance performance at 5.6%—when there were two possible choices and the choices came from different equivalence classes. However, when subjects could intervene on the structures by forcing aliens to think particular thoughts, Steyvers et al. (2003) found that performance improved to around 33%.

Although Steyvers *et al.* (2003) demonstrated that adults could distinguish between CGMs that derived from different classes based on observation and intervention, evidence was equivocal in a third experiment about whether adults were sensitive to Markov-equivalence classes: adults showed sensitivity to the Markov-equivalence classes when a common-effect structure generated the pattern of communication but showed limited sensitivity to such classes when either a common-cause or a causal-chain structure generated the communication pattern. This finding from Experiment 3 is especially interesting given that adults were explicitly shown all possible structures, including structures that comprised a Markov equivalence class, and the fact that they were unconstrained in the number of structures that they were allowed to consider. Thus, despite the fact that adults were readily able to distinguish between structures that derived from different equivalence classes, it is an open question whether adults are sensitive to the Markov condition and Markov-equivalence classes or whether, as has been demonstrated in previous research that used launching-event sequences, adults process causal events in terms of low-level perceptual features such as the spatiotemporal relation between elements and objects in a causal event. With regards to the second point, there is extensive research with infants (for a review, see Cohen et al., 1998), children (e.g., Kloos & Sloutsky, 2013; Kushnir & Gopnik, McClelland & Thompson, 2007; Rakison & Lupyan, 2008), and adults (e.g., Allan & Jenkins, 1980, 1983; Buehner & Humphreys, 2010; Scholl & Nakayama, 2002; Schlottmann & Shanks, 1992; Wasserman, Elek, Chatlosh, & Baker, 1993; Wasserman & Shaklee, 1984) that demonstrates that they rely on low-level perceptual cues such as the spatial relation or temporal contiguity between causes and effects to make causal inferences.

Finally, there is evidence that infants can encode the Markov condition in causal events. For example, using eye-tracking Sobel and Kirkham (2006) found that following exposure first to two central video events (i.e., events A and B) that always predicted a dynamic side event (event C) and then to an event in which A predicted either C (i.e., the backwards-blocking condition) or D (i.e., the indirect inference condition), 8-month-old infants at test reliably looked longer at D than C when B was cued by itself in the backwards-blocking condition but longer at the C frame than the D frame when B was cued by itself in the indirect-inference condition. Sobel and Kirkham (2006) interpreted this finding to mean that infants recognized the Markov condition in the event sequences. However, this finding should be interpreted cautiously because (a) infants' responses to the test events could have resulted either from noticing violations to the associative relations between the events or from noticing violations to the Markov condition in the BB and IS conditions, (b) the obtained pattern of looking could be interpreted in several different ways, and (b) the BB and IS conditions in the experiment bore little resemblance to these conditions used in blicket-detector studies (for a review see Shultz, 2007).

## Current Experiments

Based on the limitations of the studies described above, open and as yet unanswered questions remain about (1) whether infants and adults are sensitive to and can encode the Markov condition in causal events, (2) whether, upon encoding the Markov condition, infants and adults treat sequences that instantiate the same conditional independence relations equivalently and, as such, can form Markov-equivalence classes, and (3) how such sensitivity changes from late infancy to adulthood. In addition, it is unresolved whether, how, and to what extent, infants' and adults' sensitivity to the Markov condition is robust to low-level perceptual manipulations; that is, it is unknown whether infants’ and adults’ sensitivity to the Markov condition will hold when salient perceptual manipulations are made to causal sequences.

The research presented here consists of five experiments that used four-object, three-chain causal launching sequences. Launching sequences were used because they are among the simplest events in which to observe cause-and-effect relations between elements, because this paradigm has been used successfully to study causal perception in infants, children, and adults (e.g., Leslie & Keeble, 1987; Oakes & Cohen, 1990; Scholl & Nakayama, 2002), and because there is reason to think that adults may show greater sensitivity to the Markov condition and Markov-equivalence classes when a causal-chain structure is used than when either a common-cause or common-effect structure is used (e.g., Steyvers et al., 2003). Indeed, although Steyvers et al. (2003) demonstrated that adults showed limited sensitivity to the Markov condition and Markov-equivalence classes overall when temporal cues to causal structure were removed from causal-chain, common-cause, and common-effect structures, there is reason to think that adults will show sensitivity to the Markov condition and Markov-equivalence classes when such cues are retained, especially in the context of causal-chain sequences. Fernbach and Sloman (2009) showed, for example, that adults rely on temporal cues and use causal inference techniques such as “chaining”—in which inferences about individual causal relations are grouped together from one trial to the next—to learn causal-chain and common-cause models.

Note that the reason we chose not to consider common-effect structures here is because such sequences form their own equivalence classes—that is, no two common-effect structures instantiate the same conditional independencies. This makes it difficult to assess whether learners are sensitive to the Markov condition and can form Markov-equivalence classes between two (or more) common-effect structures because no two common-effect structures will instantiate identical independence relations. This means that if adults are exposed to one common-effect structure that implies particular causal relations during an initial training phase (e.g., Event A) and the goal is to determine whether they will form a Markov-equivalence class that includes Event A and another event at test that instantiates identical conditional-independence relations as Event A, then one of the test events must be Event A. This is problematic because if adults respond equivalently to this test event, it could be because they recognized that this test event instantiated identical conditional-independence relations as those shown during training or because this test event was simply perceived as familiar irrespective of whether it instantiated the same relations as those shown during the training phase. We chose not to use common-cause structures and instead opted to use causal-chain structures because, as was mentioned above, causal-chain structures, but not common-cause structures, are the simplest events in which to observe cause-and-effect relations.

Four object, three-chain sequences were used instead of canonical two-object, one-chain sequences (e.g., Leslie & Keeble, 1987) or instead of three-object, two-chain sequences (e.g., Cohen, Rundell, Spellman, & Cashon, 1999) because examining infants' and adults' sensitivity to multiple conditional independence relations (i.e., the Markov condition) is only possible in four-object, three-chain sequences. To our knowledge, these experiments were the first to test infants’ and adults’ causal processing with four-object, three-chain sequences. Experiment 1 examined whether adults are sensitive to the Markov condition and Markov-equivalence classes such that they will respond equivalently to test sequences that instantiate the same conditional independence relations as those shown during an initial training phase. Experiment 2 extended Experiment 1 to examine the extent to which adults’ sensitivity to the Markov condition and Markov-equivalence classes is robust to salient perceptual manipulations to the four-object launching-sequences and whether adults’ representations of such sequences are grounded in low-level perceptual features or in terms of the Markov condition and Markov-equivalence classes. Experiment 3 was designed to clarify the results of Experiment 2 to determine on what basis subjects in Experiment 2 responded to the test sequences and whether the sequences were perceived as perceptually or causally distinct. Experiment 4 examined over what dimension, or perceptual feature, do adults encode the Markov condition when shown four-object, three-chain launching event sequences. Finally, the aim of Experiment 5 was identical to that of Experiment 1 except that infants between 18 and 22 months of age were tested in a habituation version Experiment 1. Infants and adults were chosen because it is possible to examine the developmental trajectory of causal processing from infancy to adulthood.

# Experiment 1

The goal of Experiment 1 was to examine on what basis adults processed four-object, three-chain launching sequences and to assess their sensitivity to the Markov condition and Markov-equivalence classes. In particular, during an initial training phase, adults were shown four-object launching-event sequences in which preceding objects in the chain caused succeeding objects to move through immediate and direct contact. Following this initial training phase, adults were shown four test events and were asked to rate how perceptually and causally similar each test event was to the initial training event. It was predicted that adults would provide higher inconsistency ratings to the test events that violated the set of conditional independencies in the training sequence if they were sensitive to the Markov condition and Markov-equivalence classes. In contrast, it was predicted that adults would provide higher inconsistency ratings to the test events that violated the perceptual features of the training sequence if they formed a representation of the training sequence that was grounded in low-level perceptual features and cues such as which object moved first.

## Methods

**Participants.** Sixty-four college students were recruited from to participate in Experiment 1. Data about the age and sex of the subjects were not collected. An a priori power analysis revealed that approximately 16 participants would be needed to have an 80% chance of detecting a reliable difference between any two pairs of means if such a difference exists (e.g., Cohen, 1988). Thus, we can be confident that these experiments are sufficiently powered. The subjects received course credit for an introductory-level psychology course.

**Stimuli and Design.** The stimuli were computer-animated launching-event sequences that were produced using Macromedia Director MX 2004 for PC. The stage on which the video events were depicted was 641 pixels wide by 480 pixels tall. In the training phase, a first circle that was initially out of sight and off the screen moved horizontally across the stage until it contacted a second object that was located a short distance away. On contact, the first object stopped moving and the second object began horizontally to move until it contacted a third circle. Following contact with the third object, the second object stopped moving and the third object began to move horizontally until it made contact with and subsequently caused a fourth circle to move horizontally until it was off stage and out of sight. The entire training event sequence lasted approximately 10 s and each sequence was played both from left to right and from right to left and was counterbalanced for first event to eliminate potential direction effects. The colors of the first and third circles in the training sequences were either red or blue (counterbalanced), whereas the color of the second and fourth objects was always gray. Note that regardless of whether the training-event sequences were played from right to left or from left to right, the red and blue circles always served as the first and third objects (counterbalanced) and the gray objects always served as the second and fourth objects. Each circle was approximately 56 pixels wide by 56 pixels tall. Individual presentations of each event were separated by a blue screen that descended and ascended over a period of approximately 2 s.The training and test stimuli are illustrated in Figures 2A-D.

In the test phase, subjects were shown four test events: the gray-blue-gray-red (i.e., GBGR), the gray-red-gray-blue (i.e., GRGB), the gray-blue-red-gray (i.e., GBRG), and the gray-red-blue-gray (i.e., GRBG) test events. All test events were similar the training events in that they included four objects in a causal chain sequence; however, in each one the order of the objects changed from training to test. In the GBGR test event, the blue object occupied the second position, the red object occupied the fourth position, and the first and third positions were occupied by the gray circles. In the GRGB test event, the red object occupied the second position and the blue object occupied the fourth position in the sequence; the first and third positions in the sequence were occupied by the gray objects. In the GBRG test event, the blue object occupied the second position and the red object occupied the third position in the sequence; the first and fourth positions in the sequence were occupied by the gray objects. Finally, in the GRBG test event the red object occupied the second position and the blue object occupied the third position in the sequence; the first and fourth positions were occupied by the gray objects. The GBGR, GBRG, and GRBG test events were included because they enabled us to determine whether and to what extent adults processed the initial training event in terms of the independence relation between the blue and red circles; that is, whether adults formed a representation of the initial training sequence in which the blue and red circles were separated by an intermediate gray object. The GRGB test event was included because it enabled us to determine whether and to what extent adults were sensitive to the Markov condition and were sensitive to Markov-equivalence classes. Indeed, although the relative positions of the blue and red circles differed in the GRGB test event than in the initial BGRG training event, this test event nonetheless incorporated the same set of conditional independencies as that shown in the initial training event. In this way, the GRGB test event formed a Markov-equivalence class with the initial training event. The positions in the sequence occupied by the red, blue, and gray objects were counterbalanced across participants, and each test event was shown in counterbalanced order using a Latin square.

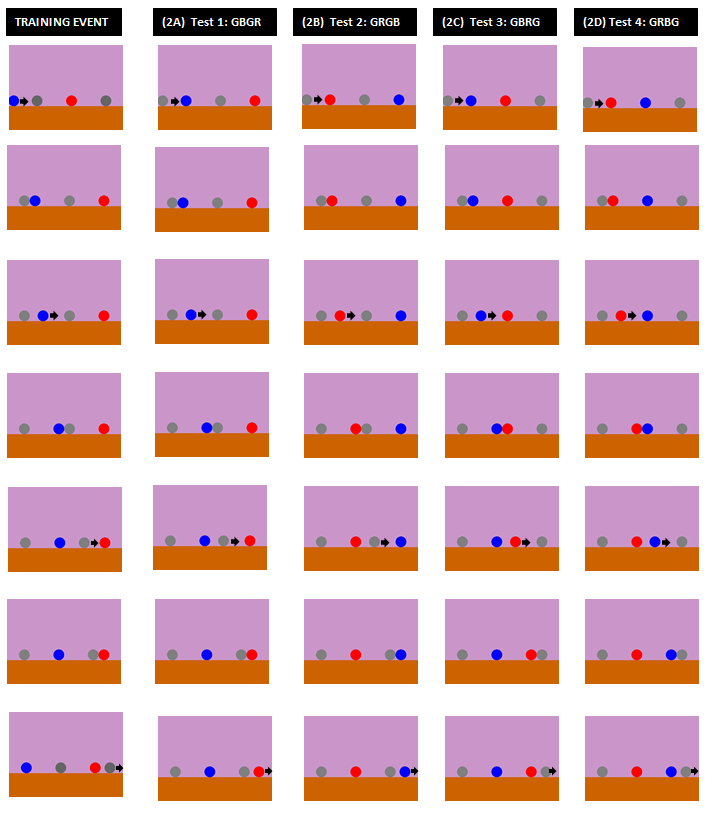


Fig. 2A-D. Examples of the training event and test events used in Experiment 1. Subjects were shown the training event three times from left-to-right and three times from right-to-left. The position of the red and blue objects was counterbalanced across subjects. Subjects then saw the four test events, and the test events were shown in counterbalanced order across subjects.

**Procedure.** During the training phase, subjects were shown each training event three times from left-to-right and from right-to-left for a total of six presentations. Following the training phase, subjects were shown each test event once in a counterbalanced order using a Latin square. Following each test event, subjects were asked to provide two different ratings. Specifically, for the perceptual question subjects were asked to “indicate how perceptually similar each test event is to the training events on scale that ranges between 0 (completely perceptually dissimilar) and 100 (completely perceptually similar).” For the causal question, subjects were asked to “indicate how causally similar each test event is to the training events on scale that ranges between 0 (completely causally dissimilar) and 100 (completely causally similar).” The rationale for requiring subjects to provide both causal and perceptual ratings was that it made it possible to determine the conditions under which adults would show sensitivity to the Markov condition based on their ratings of each test event; that is, whether adults would show sensitivity when asked to provide perceptual ratings, causal ratings, or both. Note that this would not be possible if subjects were asked only whether a particular test event was similar to the training events because the word similar could be taken either to mean perceptual similarity, causal similarity. Subjects were asked this question for each test event and were given approximately 15 s to answer. We used a 100-point scale because previous research has used a similar rating scale to assess adults' causal reasoning abilities (e.g., Benton & Rakison, 2016; Waldmann, 2001) and because the use of such a scale made it possible to determine to what extent subjects' causal and perceptual ratings of the test events were graded.

## Predictions

The test events were designed to determine on what basis subjects processed the four-object, three-chain launching-event training sequences, and whether they generalized their learned representation from training to test. The predictions for how adults should respond to the four test events are shown in Figures 5 and 6. Thus, as is shown in Figure 3 (and its inverse in Figure 6), higher causal ratings to the GBGR and GRGB test events than to the GBRG and GRBG test events would indicate that subjects perceived the GBRG and GRBG test events to be causally and/or perceptually dissimilar to the training events because the red and the blue objects were no longer separated by a gray object in these sequences and hence were independent. Higher causal and/or perceptual ratings to the second test event relative to first, third, and fourth test events would indicate that subjects had encoded all possible conditional independence relations (i.e., the Markov condition) in the training events and perceived the second test event, but not the first, third, or fourth test events, to be most similar to the training events. This is because this test event involved identical conditional independence relations as that shown during the training phase and thus formed a Markov-equivalence class with the training sequence; that is, the training event and the GRGB test event produced the following identical joint probability distribution: *P*(*B*)*P*(*G*|*B*)*P*(*R*|*G*)*P*(*G*|*R*). Note that due to space constraints we have omitted the technical details about this and other joint probability distributions discussed in this paper; the relevant mathematical details can be found in Steyvers et al. (2003).

Higher causal and/or perceptual ratings to the first and third test events relative to the second and fourth test events would indicate that they had processed the training events in terms of the temporal precedence relation between the red and blue objects; that is, they formed a representation in which the red object preceded temporally the blue object in the training sequences. Finally, high causal and/or perceptual ratings to all four test events would indicate that the adults had not encoded any particular dimension (or considered all changes to be equally important) during training but rather recognized that each test event differed from the training events in at least one of the above described ways.

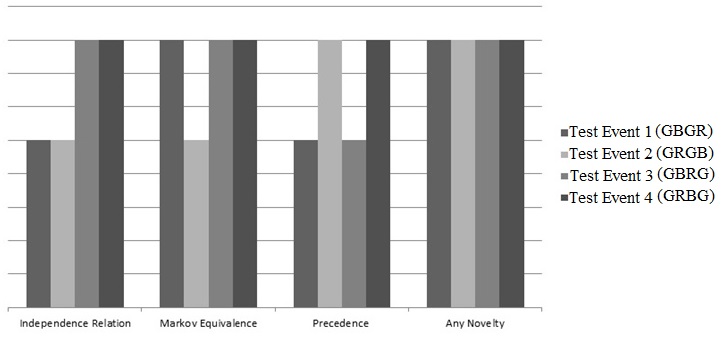


Fig. 3. Predicted pattern of responding to the test events in Experiment 1 based on whether subjects responded to the test events based on independence-relation information, the Markov condition, temporal-precedence information, or the novelty of the test events relative to the training event.

## Results

All analyses were conducted in R (R Development Core Team, 2008). Figure 4 shows the mean causal and perceptual ratings for each of the four test events. Given evidence of non-normality and unequal variance in the causal-rating data in this and all subsequent experiments, all analyses used non-parametric analyses with 4,000 replications each for hypothesis testing and to estimate confidence intervals. In addition, perceptual and causal ratings of the four test events were analyzed in this and all subsequent experiments using linear mixed-effects models (LMM). This represents a better approach than either univariate ANOVA or ordinary least squares regression because it addresses unbalanced and non-independent designs and data (for an extended discussion, see Baayen, Davidson, & Bates, 2008). Finally, all marginal differences were further analyzed using Bayes' Factor analyses.

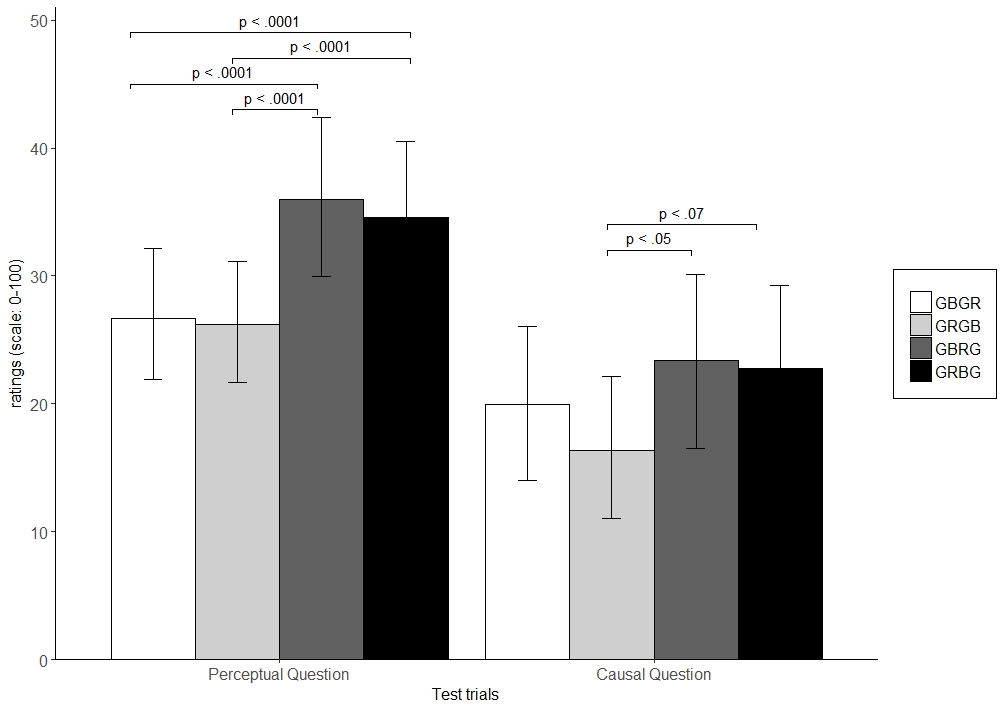


Fig. 4. Subjects’ mean perceptual and causal ratings of the four test events in Experiment 1. The black bars correspond to subjects’ perceptual ratings of the test events and the gray bars correspond to subjects’ causal ratings of the test events.

The first analysis examined whether subjects' ratings of the test events differed based on whether they were asked to provide causal or perceptual ratings first (question type: perceptual first v. causal first) and on whether the red or blue object occupied the first position in the training sequences (position: red v. blue) on which they were trained. The rationale for this analysis was to determine whether there was an effect of position or question type on subjects’ causal and perceptual ratings of the test events. Preliminary analyses revealed no main effect for question type, *F*(1, 60) = 0.18, *p* = .67, no main effect for position, *F*(1, 60) = 1.9, *p* = .16, and no significant interaction between the two variables, *F*(1, 60) = .02, *p* = .88. Consequently, subjects' ratings of the four test events were collapsed over the levels of these variables. Note that although there was no order effect of question type or location on subjects' ratings of the test events, their perceptual and causal ratings of the same events—which constituted the two levels of the explanatory variable rating type—were analyzed separately in this and all subsequent experiments reported here.

**Primary Analyses**

**Perceptual ratings.** The aim of the first main analysis was to examine whether adults' perceptual ratings of the four test events differed in their perceptual relationship to the training events. A linear mixed-effects model, with subjects as the random-effects factor and test trial as the within-subjects fixed-effects factor, revealed a significant main effect of test trial, R2= .67, *F*(3, 189) = 7.26, *p* < .0001. Note that because our aim was to determine which events subjects perceived as inconsistent with the training sequence, subjects’ perceptual and causal ratings were subtracted from 100 to obtain inconsistency ratings that reflected the extent to which a particular test event was perceived as inconsistent with the training event rather than as consistent with it. Note that using subjects' adjusted ratings instead of their raw ratings did not affect any of the analyses reported here given that the distribution of responses and the direction of the analyses remained the same regardless whether subjects' raw ratings or adjusted ratings were used.

Follow-up planned comparisons using permutation testing indicated that perceptual ratings of the GBGR event (*M* = 26.67, Bootstrapped 95% CI[21.55, 31.79]) did not differ from ratings of the GRGB event (*M* = 26.19, Bootstrapped 95% CI[21.15, 31.22]), *p* = .92. In other words, the GBGR and GRGB test events were perceived as equally consistent with the training event. However, subjects provided higher inconsistency ratings of the GBRG (M = 35.95, Bootstrapped 95% CI[29.94, 41.97]) and GRBG (*M* = 34.61, Bootstrapped 95% CI[28.76, 40.46]) tests events than either the GBGR test event or the GRGB test event, all *p*'s < .0001. This indicates that the GBRG and GRBG test events were perceived as more inconsistent with the training events than either the GBGR or GRGB test events. Together, these analyses indicate that subjects’ representation of the training events was based on the independence relationship between the red and blue objects; that is, the test events in which the red and blue objects were spatially and temporally contiguous were perceived as more inconsistent than the test events in which the two objects were independent presumably because these objects were separated spatially in the training sequence. Thus, when asked to make perceptual comparisons, adults’ representations of the causal events used here appear grounded in low-level perceptual features.

**Causal ratings.** The second main analysis examined adults’ causal ratings of the four test events. A linear mixed-effects model, with subjects as the random-effects factor and test trial as the within-subjects fixed-effects factor, revealed a significant main effect of test trial, R2 = .8, *F*(3, 189) = 3.6, *p* < .025. Follow-up planned comparisons using permutation testing indicated that subjects provided higher inconsistency causal ratings of the GBGR event (*M* = 19.94, Bootstrapped 95% CI[13.79, 26.08]) than of the GRGB event (*M* = 16.39, Bootstrapped 95% CI[10.74, 22.04]), *p* < .0001. In contrast, subjects ratings of the GBGR event did not differ from their ratings of either the GBRG event (*M* = 23.36, Bootstrapped 95% CI[16.63, 30.08]) or the GRBG event (*M* = 22.79, Bootstrapped 95% CI[16.18, 29.42]), both *p*'s > .2. However, subjects provided higher causal ratings of the GBRG event than the GRGB event, *p* = .05, and marginally higher ratings of the GRBG test event than the GRGB event, *p* = .07.

To determine whether both comparisons (i.e., the GBRG v. GRGB comparison and the GRBG v. GRGB comparison) were meaningful and provided evidence in support of the alternative hypothesis that the pairs of mean causal ratings for both comparisons differed, Bayes’ Factors were estimated using Bayesian Information Criteria (Wagenmakers, 2007). In this analysis, the fit of the data under the null hypothesis (i.e., no difference between the two test events) was compared to that under the alternative hypothesis (a two-tailed difference between the ratings). The first analysis revealed that the observed difference between the causal ratings of the GRBG and the GRGB test events (∆GRBG-GRGB = -6.41) was 7.86 times more likely under the alternative hypothesis than the null hypothesis. Using Raftery (1995) and Jeffrey'(1961) guidelines for interpreting Bayes’ Factors, a Bayes’ Factor of 7.86 represents positive and substantial evidence in support of the alternative hypothesis. The second analysis revealed that the observed difference between the causal ratings of the GRGB and the GBRG test events (∆GRGB-GBRG = -6.97) was 20.82 times more likely under the alternative hypothesis than the null hypothesis. Using both Raftery (1995) and Jeffrey'(1961) guidelines, this represent strong evidence in support of the alternative hypothesis. Thus, the differences between the causal ratings of the GRGB and GRBG and the GBRG and GRGB test events represent reliably meaningful differences.

The results from subjects' causal ratings of the four test events support the Markov condition predictions and suggest that adults were sensitive to the Markov-equivalence class formed between the training event and the GRGB test event only when they were asked to rate how causally similar each test event was to the training sequence.

**Individual Differences**

Additional analyses were undertaken in this and all subsequent experiments because it is possible that the distribution of responses to the perceptual and causal questions in Figure 1 could have masked important individual differences in behavior. These analyses were also conducted because our aim was to determine whether the proportion of subjects who showed sensitivity to the Markov condition and Markov-equivalence classes differed from that of subjects who responded to the test events based on lower-level perceptual cues. Figure 5 shows the inconsistency ratings of the four test events for the perceptual question and the causal question for each of the 64 total subjects in the experiment.

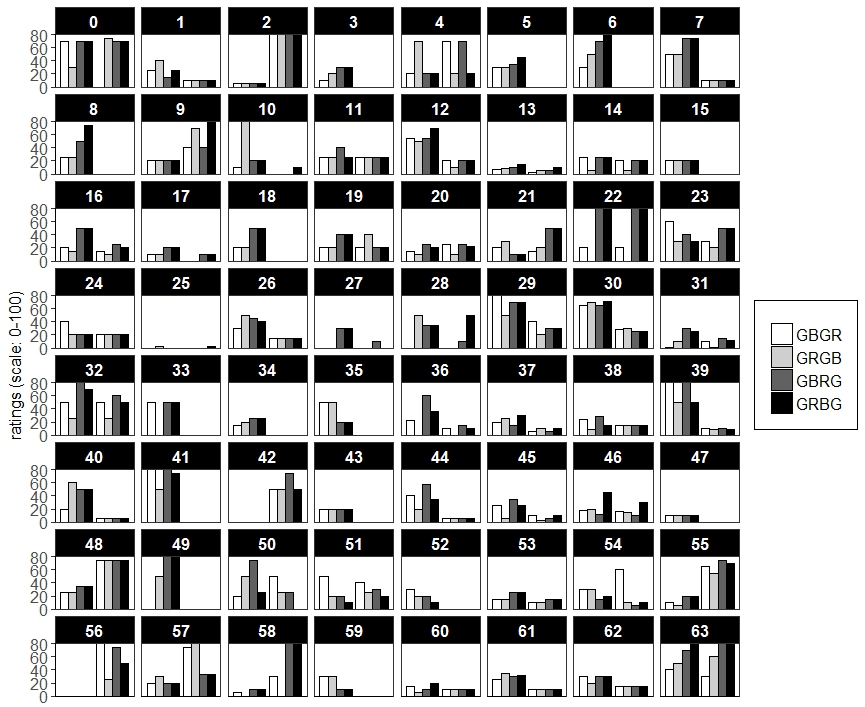


Fig. 5. Each of the 64 subjects’ causal and perceptual ratings of the four test events in Experiment 1.

Note that because attention to the independence or temporal-precedence relation between the blue and red objects indicates sensitivity to low-level perceptual features and cues, subjects whose distributions suggested sensitivity to these relations were classified as Perceptual processors. Likewise, subjects whose inconsistency ratings were either at ceiling or floor for all four test events—indicating that all four test events were perceived as novel or familiar, respectively—were counted as Perceptual processors. This is because if subjects represented the training events as launching events in the most general sense (i.e., as a general launching event without recourse to the perceptual differences between particular launching events), then their inconsistency ratings of all four test events should be at floor because all four test events represent launching events. In contrast, subjects should provide inconsistency ratings at ceiling for all four test events if they recognized that each test event differed from the training events in at least one way. Nonetheless, subjects’ whose inconsistency ratings were at ceiling were included in the Perceptual processors group. In contrast, subjects whose distribution of responses to the test events suggested sensitivity to the Markov condition and Markov-equivalence classes formed the "Markov" group; that is, subjects who provided higher inconsistency ratings to the GBGR, GBRG, and GRBG test events, but not to the GRGB test event—and thus whose pattern of responses suggested sensitivity to Markov-equivalence classes—comprised the Markov group. Finally, subjects whose responses to the test events neither suggested sensitivity to Markov-equivalence classes nor to low-level perceptual cues formed the Other group. It is worth mentioning that although it was possible to provide a more nuanced analysis that compared the proportion of subjects who attended to any one of the low-level perceptual cues to each other and to that for participants who encoded the Markov condition, statistical power was increased by combining participants into either the Perceptual, the Markov, or the Other group.

An analysis of individual differences for the perceptual question revealed that a larger proportion of subjects were classified as Perceptual processors (N = 34; 53.125%) than as "Markov" processors (N = 18; 28.125%) or as Other processors (N = 12; 18.75%), binomial tests, both *p*’s < .0001. Of the 34 subjects who were classified as Perceptual processors, 23 subjects processed the causal events in terms of the independence relation between the blue and red circles; that is, subjects’ responses implied that they formed a representation in which the blue and red circles were separated by a gray circle and perceived as more inconsistent those test events in which the blue and red circles were spatially contiguous. This pattern of responding is consistent with that discussed above in the Main Analysis section and that shown in Figure 6. The proportion of subjects who were classified as Markov processors did not differ from Other processors, however, binomial test, *p* = .36.

Likewise, for the causal question, a larger proportion of subjects were classified as Perceptual processors (N = 40; 62.5%) than as Markov processors (N = 15; 23.44%) or Other processors (N = 9; 14.06%), binomial tests, both *p*’s < .005. Of the 40 subjects who were classified as Perceptual processors, 25 subjects processed the causal events in terms of familiarity information; that is, their inconsistency ratings of all four test events were at floor, which suggests that all four test events may have been perceived as familiar because they were different instantiations of more general launching-event sequences. However, the proportion of subjects who were classified as Markov processors did not differ from Other processors, binomial test, *p* = .31.

## Discussion

Experiment 1 was designed to determine on what basis adults process complex launching-event sequences that are composed of four objects, and whether adults could form Markov-equivalence classes by responding equivalently to structures that instantiated the same full set of conditional independencies (i.e., the Markov condition). The overall pattern of results demonstrated that adults assigned higher inconsistency ratings to the test events that violated the Markov condition seen during training when asked to compare to what extent a particular test event was causally similar to the training events. However, when asked to compare to what extent a particular test event was perceptually similar to the training events, they assigned higher inconsistency ratings to the test events in which red and blue objects were no longer separated by an intermediate object and were spatially contiguous. This finding is interesting because it suggests that adults represent complex causal events at multiple levels and that the aspect of the representation that is accessed depends crucially on whether they are asked to evaluate and compare events on the basis of their causality or perceptual similarity. This finding is also important because this is the first time that it has been demonstrated that adults can form Markov-equivalence classes in the context of causal-chain sequences that are composed of four objects. This means that temporal cues to causal structure need not be removed—as has been suggested in previous research (e.g., Steyvers *et al.*, 2003)—for adults to be sensitive to Markov-equivalence classes.

In addition to this overall finding, an analysis of individual differences revealed that a much larger proportion of adults represented the causal events used here in terms of low-level perceptual cues and features than in terms of the Markov condition, regardless of whether they were asked to make causal or perceptual comparisons. This indicates that although adults show sensitivity to the Markov condition and Markov-equivalence classes when shown complex launching sequences that are composed of four objects, they have a greater tendency to process these events in terms of low-level perceptual features than in terms of the Markov condition. From a theoretical perspective, this result suggests that sensitivity to the Markov condition may not be as robust or universal as has been thought, especially when analyzed at the level of individual subjects (e.g., Gopnik *et al.*, 2004; Steyvers *et al.*, 2003). Nonetheless, the results from Experiment 1 indicate that adults can respond to complex launching sequences that are composed of four objects in terms of low-level perceptual features and the Markov condition and Markov-equivalence classes, although the ability to represent the Markov condition and Markov-equivalence classes is restricted.

One unresolved question from Experiment 1 is why adults responded in two different ways when asked to make causal comparisons and perceptual comparisons. One possibility is that they may have erroneously surmised that it was a requirement to respond to the test events in two different ways. This may have been based on the fact that they were asked two distinct questions that used distinct adjectives (i.e., *causal* vs *perceptual*). However, this possibility is unlikely because such an account makes no discernible prediction about how subjects’ ratings should be distributed across the four test events for both questions. There is also no reason to expect the pattern of results in response to the causal and perceptual questions that were observed here if subjects felt compelled to respond to the test events in two different ways. This is because there were a relatively large number of potential response patterns that could have been observed when asked the two questions. This account predicts only that adults' inconsistency ratings should be randomly distributed with the sole constraint being that the distribution of responses to the four test events for the causal question be different than that for the perceptual question.

A more likely possibility is that the two questions tapped two different representations (or different aspects of the same representation) against which each of the four test events were compared. To examine this possibility, the aim of Experiment 2 was conceptually to replicate the overall finding from Experiment 1 in a modified version of the task. In particular, we examined whether adults would, as in Experiment 1, show sensitivity to the Markov condition when asked to make causal comparisons but not when asked to make perceptual comparisons. The rationale for this decision was to assess the reliability of this finding from Experiment 1. Finally, given that a non-negligible proportion of subjects showed sensitivity to the Markov condition and Markov-equivalence classes in Experiment 1, we also examined in Experiment 2 to what extent this sensitivity is robust to salient perceptual manipulations such as the inclusion of a spatial gap in four-object, three-chain launching event sequences.

# Experiment 2

Experiment 2 was designed to examine whether adults in a modified version of the task used in Experiment 1 would show differential sensitivity to the Markov condition and Markov-equivalence classes when asked to make two different comparisons. Experiment 2 was also designed to examine to what extent adults' sensitivity to the Markov condition was robust to salient perceptual manipulations. This is an important issue to examine because a number of proponents of CGM accounts (e.g., Gopnik et al., 2001; Waldmann, 1996) maintain that so long as such salient perceptual manipulations do not affect the underlying set of conditional independencies (i.e., the Markov condition), learners should show sensitivity to the Markov condition and Markov-equivalence classes. The training events used in Experiment 2 were identical to those used in Experiment 1. However, Experiment 2 differed from Experiment 1 in terms of the test events that were used: two of the four test events incorporated a spatial gap between two objects, whereas the remaining two test events did not incorporate a spatial gap. Critically, one of the gap events and one of the non-gap events encapsulated the same conditional independencies as those shown during training. Thus, if adults were sensitive to the Markov condition in the training event, they should perceive the gap and non-gap test events that incorporated the same conditional independencies as most similar to the training event. However, if adults’ sensitivity to the Markov condition was not robust to the inclusion of a spatial gap, then they should perceive the non-gap events as most similar to the training event. This is despite the fact that one of the non-gap events did not incorporate the same conditional independencies as the training event.

## Methods

**Participants.** Sixty-four college students were recruited from Carnegie Mellon University to participate in Experiment 2. Data about the age or sex of the subjects were not collected.

**Stimuli and Design.** Similar to Experiment 1, the stimuli in Experiment 2 were computer-animated four-object, three-chain launching-event sequences and the stage on which the events were depicted was the same as that in Experiment 1. The two training sequences used in Experiment 2 were identical to those used in Experiment 1 and as in Experiment 1 both sequences were played from right to left and from left to right over approximately a 10 s period. Individual presentations of each event were similarly separated by a blue screen that descended and ascended over a period of approximately 2 s.

With the exception of the first test event in which the blue (or red) object occupied the second position, the red (or blue) object the fourth position (i.e., the GBGR event), and the first and third positions occupied by the gray circles, the remaining three test events differed from those used in Experiment 1. In the second test event (i.e., the GBgapGR event), a first gray circle, which was initially out of sight and off the screen, moved horizontally across the stage until it contacted a second blue (or red) object that was located a short distance away. On contact, the first object stopped moving and the second object began horizontally to move until it was approximately 40 pixels away from the third object, at which point it stopped moving. When the second object stopped moving, a second gray object began without contact to move horizontally until it made contact with and subsequently caused a final red (or blue) circle to move horizontally until it was off stage and out of sight. The third test event was similar to the first test event except that the positions of the final red (or blue) and gray objects were switched in the third test event relative to the first test event (i.e., the GBRG event). Thus, in this sequence, the red and blue circles were spatially contiguous. The fourth test event was similar to third test event except that the spatial gap was inserted between the red and blue circles (i.e., the GBgapRG event). Thus, rather than the blue (or red) circle causing the red (or blue) circle to move through immediate and direct contact, the blue circle stopped approximately 40 pixels short of the red circle and the red circle began unaided to move. Note that all four test events were presented in counterbalanced order using a Latin square.

## Predictions

It was predicted that if subjects encoded the local independence relation between the red and blue circles during the training phase and formed a representation in which the two objects were seen as being separated by a gray object, then should assign higher inconsistency ratings to the third test event. This is because the red and blue circles were spatially contiguous—and hence unconditionally dependent—in this sequence. However, if subjects encoded the full Markov structure between each of the elements during the training phase—and were sensitive to the fact that the red circle was independent of the blue circle conditional on the gray circle and that the two gray circles were independent of each other conditional on the red circle (i.e., to the full set of conditional independencies)—then they should provide higher inconsistency ratings to the first three test events. This is because the independence relations between the four objects in these sequences violated that seen during initial training.

To understand why subjects should not assign higher inconsistency ratings to the fourth test event if they encoded the Markov condition in the training sequences, consider the joint-probability distributions (JPD) for each structure.

For the training sequence, the JPD is as follows:

*P*(*B*)*P*(*G*|*B*)*P*(*R*|*G*)*P*(*G*|*R*)

For the fourth test event, the JPD is as follows:

*P*(*B*)*P*(*G*|*B*)*P*(*R*|*G*)*P*(*G*|*R*)

That both test events produce identical JPDs indicates that the fourth test event forms a Markov-equivalence class with the training sequence, despite the fact that a salient perceptual gap is included in this event. In contrast, if subjects were sensitive to the fact that the objects in the training sequences only moved upon contact from other objects (i.e., were spatially contiguous), then they should provide higher inconsistency ratings to the second and fourth test events. This is because these two events introduce a novel spatial gap between two of the circles that, critically, was not shown during the initial training phase.

If subjects represented the training events as launching events in the most general sense, then their inconsistency ratings of all four test events should be at floor because all four test events represent general instances of launching events. In contrast, subjects should provide inconsistency ratings at ceiling for all four test events if they recognized that each test event differed from the training events in at fewest one way. Finally, if subjects were sensitive to the fact that the blue and red objects were not spatially contiguous in the sequence, they should assign higher inconsistency ratings to the third and fourth test events. This is because in these sequences the blue and red objects were spatially contiguous. The predictions are displayed below in Figure 6.

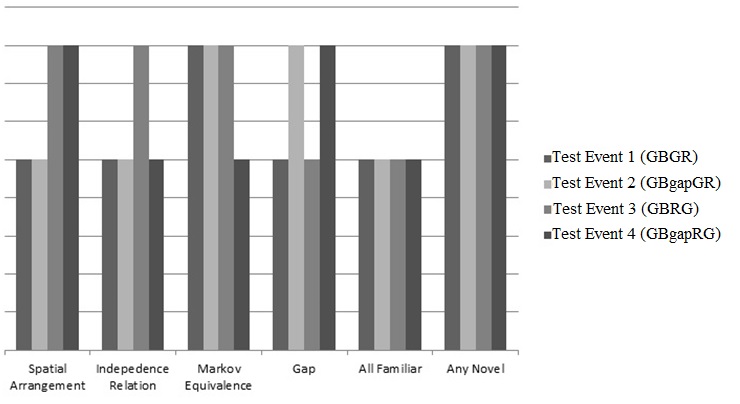


Fig. 6. Predicted pattern of responding to the test events in Experiment 2 based on whether subjects responded to the test events based on differences relative to the training event in six features.

**Procedure.** The procedure for Experiment 2 was identical to that for Experiment 1; that is, adults were shown the left-to-right training event three times and the right-to-left training event three times. Following the training phase, subjects were shown each test event once in counterbalanced order using a Latin square.

## Results

The first analysis examined whether subjects' ratings of the test events differed based on whether they were asked first to provide causal or perceptual ratings (question type: perceptual v. causal) and on whether the red or blue object occupied the first position in the training sequences (position: red v. blue) on which they were trained. A linear mixed-effects model indicated a marginally significant effect for question type, *F*(1, 446) = 3.75, *p* = .05, and a significant interaction between position and question type, *F*(1, 446) = 6.74, *p* < .01. However, the main effect of position was not significant, *F*(1, 62) = .003, *p* = .95. Post-hoc permutation tests revealed that subjects who were shown training sequences in which the red object was the first object in the sequence did not differ in their inconsistency ratings when asked to make causal comparisons (*M* = 33.19, Bootstrapped 95% CI[28.83, 37.56]) than when asked to make perceptual comparisons (*M* = 33.5, Bootstrapped 95% CI[29.18, 37.82]), *p* = .94. However, subjects who were shown training sequences in which the blue object was the first object in the sequence provided higher inconsistency ratings when asked to make causal comparisons (*M* = 38.81, Bootstrapped 95% CI[33.86, 43.76]) than when asked to make perceptual comparisons (*M* = 27.48, Bootstrapped 95% CI[22.66, 32.29]), *p* < .001.

Although it is unclear why this might have been the case, this finding is nonetheless consistent with the fact that in general subjects provided higher inconsistency ratings when asked to make causal comparisons than when asked to make perceptual comparisons. This was evidenced by the marginally significant main effect for question type. One reason adults may have provided higher inconsistency ratings when asked to make causal comparisons than when asked to make perceptual comparisons is that the causal question may have been easier to interpret than the perceptual question. This, in turn, could have caused subjects to feel more confident in their assignment of inconsistency ratings to the four test events—and thus may have contributed to higher inconsistency ratings—than when asked to make perceptual comparisons.

**Primary Analysis**

**Perceptual ratings.** Figure 7 shows the mean causal and perceptual ratings for each of the four test events. The aim of the first primary analysis was to examine whether adults' perceptual ratings of the four test events differed in their perceptual relationship to the training sequence. A linear mixed-effects model, with subjects as the random-effects factor and test trial as the within-subjects fixed-effects factor, revealed a significant main effect of test trial, R2 = .72, *F*(3, 189) = 55.51, *p* < .0001.Follow-up planned comparisons using permutation tests indicated that subjects provided higher inconsistency ratings to the GBgapGR (*M* = 49.55, Bootstrapped 95% CI[43.74, 55.35]) and GBgapRG (*M* = 47.63, Bootstrapped 95% CI[42.26, 52.99]) test events than to the GBGR test event (*M* = 22.34, Bootstrapped 95% CI[16.34, 28.35]) or the GBRG test event (*M* = 22.03, Bootstrapped 95% CI[16.33, 27.73]), all *p*'s < .0001. In contrast, subjects inconsistency ratings of the GBGR test event did not differ from their ratings of the GBRG test event, *p* = 1. Likewise, subjects' inconsistency ratings of the GBgapGR test event did not differ from their ratings of the GBgapRG test event, *p* = 1.

**Causal ratings.** The aim of the second analysis was to examine whether subjects' causal ratings of the four test events differed. A second mixed-effects linear model, with subjects as the random-effects factor and test trial as the within-subjects fixed-effects factor, revealed a significant main effect of test trial, R2 = .65, *F*(3, 189) = 59.33, *p* < .0001. Follow-up planned comparisons using permutation testing for the causal question indicated that subjects provided higher inconsistency ratings to the GBgapGR (*M* = 49.03, Bootstrapped 95% CI[43.31, 54.75]) and GBgapRG (*M* =42.84, Bootstrapped 95% CI[36.81, 48.87]) test events than to the GBGR (*M* = 17.38, Bootstrapped 95% CI[11.92, 22.83]) or GBRG test events (*M* = 15.34, Bootstrapped 95% CI[ 10.73, 19.96]), all *p*'s < .0001. In contrast, subjects' inconsistency ratings of the GBGR test event did not differ reliably from their ratings of the GBRG, *p* =.32. Finally, subjects' ratings of the GBgapGR test event did not differ reliably from their ratings of the GBgapRG test event, *p* = .09. These results suggest that adults were not sensitive to the Markov condition and did not form Markov-equivalence classes.

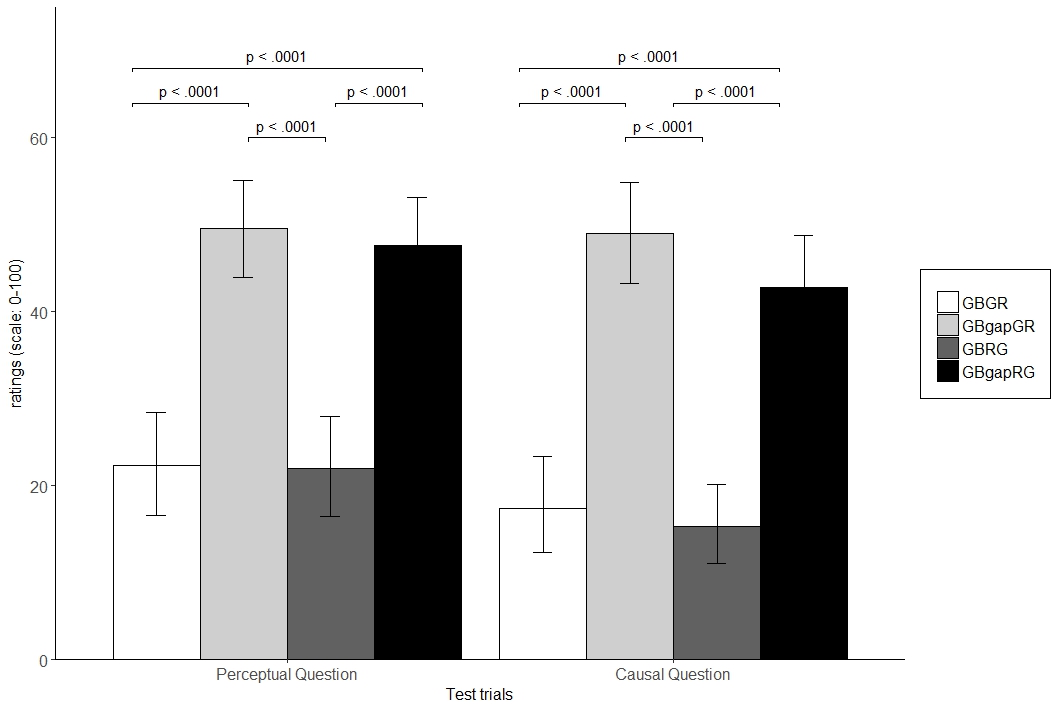


Fig. 7. Subjects’ perceptual and causal ratings of the four test events in Experiment 2. The black bars correspond to subjects’ perceptual ratings of the test events and the gray bars correspond to subjects’ causal ratings of the test events.

**Individual Differences**

The aim of this section was to examine individual differences in inconsistency ratings of the test events to determine what proportion of subjects could be classified as Markov processors, Perceptual processors, or Other processors. Subjects were classified in a similar way to Experiment 1 expect that Perceptual processors in this experiment also included subjects who showed sensitivity to the spatial gap. Figure 8 shows the inconsistency ratings of the four test events for the perceptual question and the causal question for each of the 64 total subjects in the experiment.

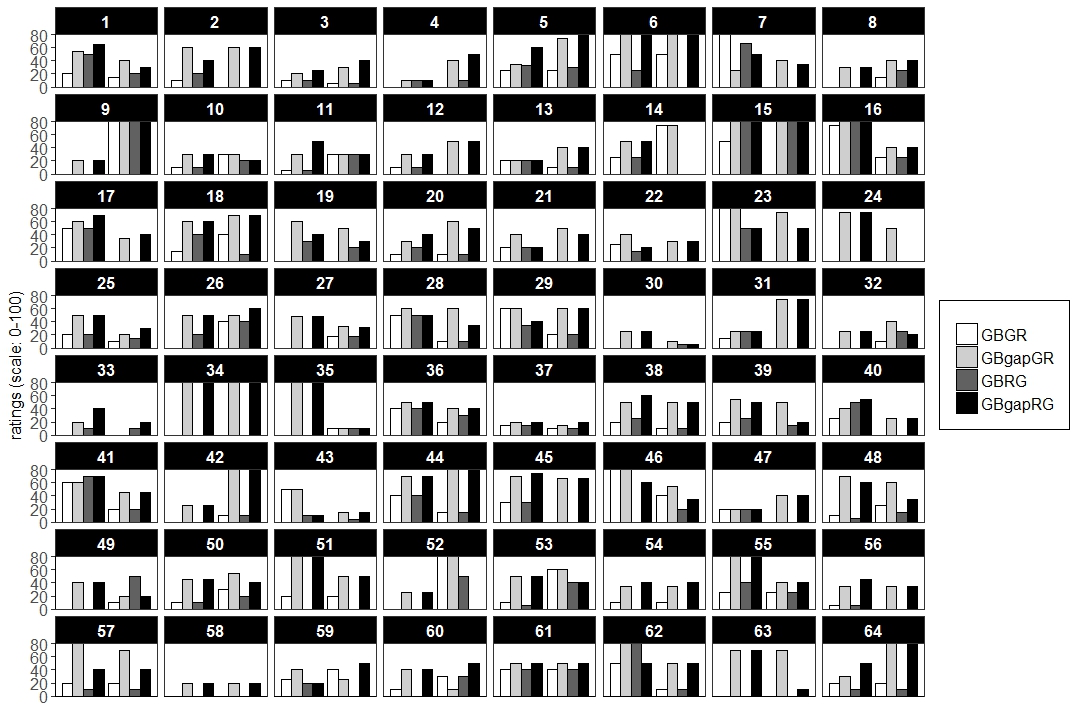


Fig. 8. Each of the 64 subjects’ causal and perceptual ratings of the four test events in Experiment 2.

An analysis of individual differences for the perceptual question revealed that a larger proportion of subjects were classified as Perceptual processors (N = 51; 79.69%) than as "Markov" processors (N = 0; 0%) or as Other processors (N = 13; 20.31%), binomial tests, both *p*’s < .00001. Of the 51 subjects who were classified as Perceptual processors, 47 processed the events in terms of spatial contiguity information; that is, subjects were sensitive to the fact that the objects in the training sequence only moved upon contact from other objects (i.e., were spatially contiguous) and deemed the test events that incorporated a spatial gap to be inconsistent with the training sequence. Moreover, a larger proportion of subjects were classified as Other processors than as Markov processors, binomial test, *p* = .001.

Likewise, for the causal question, a larger proportion of subjects were classified as Perceptual processors (N = 53; 82.81%) than as Markov processors (N = 1; 1.56%) or Other processors (N = 10; 15.63%), binomial tests, both *p*’s < .00001. Likewise, a larger proportion of subjects were classified as Other processors than Markov processors, binomial test, *p* = .0275. Similarly, of the 53 subjects who were classified as Perceptual processors, 48 subjects responded in a manner that suggested that they were sensitive to the fact that contact was required to make the objects move in the training sequence.

## Discussion

The results from Experiment 2 indicated that regardless of whether adults were asked to make causal or perceptual comparisons, they provided higher inconsistency ratings to the test events that incorporated a spatial gap.

This finding is important for two reasons. First, it failed to replicate the overall finding from Experiment 1 in which adults provided higher inconsistency ratings to the test events that violated the Markov condition and the independence relation shown in the training sequence when asked to make causal and perceptual comparisons. In Experiment 2 adults provided higher inconsistency ratings to the test events that introduced a spatial gap regardless of the comparison they were asked to make. Note again that the reason adults should not have provided higher inconsistency ratings to the last event, the GBgapRG test event, was because this event instantiated the same set of conditional independencies—despite the presence of the spatial gap—as those in the training sequence. As such, the training event formed a Markov-equivalence class with the fourth test event.

Second, this finding indicates neither that adults represent four-object launching sequences as CGMs—because they did not show evidence that they were sensitive to the Markov condition and Markov-equivalence classes—nor that their sensitivity to the Markov condition and Markov-equivalence classes is robust to salient perceptual manipulations. An analysis of individual differences also corroborated this finding: it was observed that 79.69% and 82.81% of subjects provided higher inconsistency ratings to the test events that incorporated a spatial gap when asked the perceptual and causal question, respectively, than to the test events that violated any of the other perceptual dimensions or Markov condition. Recall again that to the extent that salient perceptual manipulations do not alter the set of conditional independencies underlying a causal event, proponents of CGM accounts maintain that adults should show sensitivity to the Markov condition and be able to form Markov-equivalence classes with structures that instantiate the same set of conditional independencies (e.g., Gopnik & Tenenbaum, 2007).

These findings are important because they are the first to demonstrate that adults' sensitivity to the Markov condition and Markov-equivalence classes is not sufficiently robust to hold for perceptual manipulations to a launching event. This finding cannot be attributed to the fact that adults simply do not show sensitivity to Markov-equivalence classes in launching event sequences—as Steyvers *et al*. (2003) suggested—given the data from Experiment 1 indicated that adults are sensitive to Markov-equivalence classes in four-object launching event sequences.

That the test events that incorporated a spatial gap were perceived as more inconsistent with the training event than any other test event—including test events that violated the Markov condition in the train sequences—suggests (1) that adults’ representation of complex causal events may be grounded in low-level perceptual features and cues although there was some evidence in Experiments 1 and 2 that adults were sensitive to the Markov condition and Markov-equivalence classes, and (2) that adults’ representations of causal launching events include information about whether or not objects are made to move in the absence of contact (i.e., move in the presence of spatial gaps). Although subjects in Experiment 1 relied on different cues to represent the training and test sequences than subjects in Experiment 2, that these cues were perceptual—and possibly associative (discussed in the General Discussion)—in nature demonstrate that they may process complex launching events such as those used here perceptually rather than in terms of the Markov condition. Together, Experiments 1 and 2 are the first to illustrate that adults’ sensitivity to the Markov condition and Markov-equivalence classes depends on how adults are asked to evaluate complex causal events and whether such events include salient perceptual manipulations. That adults’ sensitivity to the Markov condition and Markov-equivalence classes did not hold in the presence of a spatial gap suggests that the overall finding in Experiment 1 was driven by the subset of subjects who were classified as Markov processors.

One unresolved question from Experiment 2 concerns on what basis adults assigned higher inconsistency ratings to the test events that incorporated a spatial gap; that is, it is unknown whether adults assigned higher inconsistency ratings to the test events that incorporated a spatial gap because these events were perceptually distinct from the training event or because they were causally distinct from the training event. If, for example, adults perceived the training event as a causal event—in which each object (other than the last one) causes other objects immediately to move through direct contact—then they should have provided higher inconsistency ratings to the test events that included a spatial gap than to the test events that did not incorporate such a gap because these events represent non-causal events. Despite it being well-established that humans treat causal launching events differently than non-causal events (Choi & Scholl, 2006; Leslie, 1984; Morris & Peng, 1994; Oakes & Cohen, 1990), an unanswered question is whether adults in Experiment 2 perceived the test events that incorporated a spatial gap as causally distinct or perceptually distinct from the training sequence. We sought to clarify the interpretation of these results in Experiment 3.

# Experiment 3

Experiment 3 we examined whether adults assigned higher inconsistency ratings to the test events that included a spatial gap because these test events were non-causal relative to the training event or because these test events were perceptually distinct from the training event. Experiment 3 used launching events like Experiments 1 and 2. However, unlike Experiments 1 and 2 that included two perceptually distinct and salient circles (namely, the red and blue circles), in Experiment 3 only one of the objects in the training and test sequences was perceptually distinct and salient (i.e., a green circle). The nature of the test events in Experiment 3 also differed from those in Experiments 1 and 2: The test events in Experiment 3 were designed to manipulate independently causality and perceptual saliency (this issue is discussed in detail below).

## Methods

**Subjects.** Thirty-two college students were recruited from Carnegie Mellon University to participate in Experiment 3. There were approximately equal numbers of males and females in the final sample. Data from one additional participant was excluded for failing to understand the directions of the task. An a priori power analysis revealed that approximately 16 participants would be needed to have an 80% chance of detecting a reliable difference between any two pairs of means if such a difference exists (e.g., Cohen, 1988). Therefore, this sample size was considered sufficient.

**Stimuli.** Similar to Experiments 1 and 2, the stimuli in Experiment 3 were computer-animated launching-event sequences and the stage on which the events were depicted was the same as that in Experiments 1 and 2. The dimensions of the circles used in Experiment 3 were also identical to those used in Experiments 1 and 2. However, the train and test sequences used in Experiment 3 differed from those used in Experiments 1 and 2. During training, subjects were shown a four-object launching sequence in which a first object that was gray made contact with and subsequently caused a second object that was green to move. The green object continued to move until it made contact with a third object that was gray, which then began to move until it made contact with a fourth object that was also gray (which then moved until off screen).

Critically, and unlike Experiments 1 and 2, the first object in the train sequences began to move from on stage rather than off stage. The rationale for this change was that it would not be possible to manipulate perceptual saliency (i.e., the presence of a spatial gap) or causality (i.e., whether or not the causal role of the salient object is preserved across learning phases) separately if the first object began off rather than on stage. To understand why, consider a hypothetical four-object launching-event sequence such as A1🡪B🡪A2🡪A3, where A1 begins to move from out of sight off stage, where B is the only salient cue, where all A objects are perceptually identical, and where the arrows indicate direction of causality. In this sequence, B is caused to move (by its parent A1). However, because A1 begins to move from offstage rather than onstage, it is impossible to know whether A1 was caused-to-move or began to move on its own and was self-propelled; that is, by starting offstage, it is impossible to know the causal role of A1 as either an object that was caused to move or that initiated its own movement. This means that simply changing B's causal status by swapping its position with A1—in an effort, for example, to change its causal status from a caused-to-move object to a self-propelled one—would be an insufficient manipulation. This is because it is unclear to what extent A1 is self-propelled or caused-to-move. Because "causality," as it is used here, refers to the causal role of the single perceptually salient object, this means that in the present experiment a test event will still be considered causally familiar with a training event if the causal role of the green circle remains unchanged between training and test, regardless of whether the causal roles of the gray objects are held constant.

Following exposure to the training four-object launching event, subjects were shown four test events: GNCN, GNCF, GFCN, GFCF. In the GNCN (i.e., Gap-Novel; Causal-Novel) test event, a gap was inserted between the first object and the second green object but not between the green object and a third gray object. By inserting a gap between the first and second object, the causal role of the second object is changed relative to that seen during training; that is, inserting the gap causes the green circle to become self-propelled, whereas during training—in which no spatial gap is present—the green circle was caused-to-move. In the GNCF test event, a gap is inserted between the green object and the third gray object, but not between the first gray object and the second green object. This event differs from those seen during training in terms of the presence of a spatial gap in which no such gap exists during training. However, the causal role of the green object remains unchanged between the training phase and this test event: in both phases, it is caused-to-move. In the GFCN test event, the position of the first and second object is reversed and, as such, the causal role of the green object is changed between the initial training event and this test event. No such spatial gap is included in this test event. Finally, the GFCF test event is identical to the training event; that is, gap presence and causality is preserved between training and test for this event. As was mentioned earlier, these test events enabled independent manipulation of gap presence (i.e., perceptual saliency) and causality (i.e., caused-to-move vs. self-propelled). In this way, these test events made it possible to establish on what basis adults assign higher inconsistency ratings to test events that incorporate a spatial gap and, as such, have the potential to clarify the results of Experiment 2.

## Predictions

If adults represented the training sequence as a direct-launching event and recognized that the green object was caused-to-move in this sequence, then they should provide higher inconsistency ratings to the GFCN and the GNCN test events. This is because the causal role of the green circle was changed in these test sequences relative to that seen during initial training. In contrast, if adults attended to the perceptual features of the train event and recognized that the train sequence did not include a spatial gap, then they should provide higher inconsistency ratings to the GNCF and GNCN test events. This is because these two test sequences introduce a spatial gap, which was not shown during the initial training phase. In contrast, if the adults encoded the Markov condition in the training sequence, then they should provide higher inconsistency ratings to the GFCN and GNCF test events because the full set of conditional independencies in these sequences differed from that in the train sequence. In addition, if adults responded on the basis of any difference, then they should assign higher inconsistency ratings to the GFCN, GNCF, and GNCN test events because each of these test events introduced at least one perceptual difference relative to the training phase. Finally, if subjects represented the training events as launching events in the most general sense, then their inconsistency ratings of all four test events should be at floor. This is because all four test events represent different instances of launching events. These predictions are shown in Figure 9.

Fig. 9. Predicted pattern of responding to the test events in Experiment 3 based on differences relative to the training phase in the causal role of the green object, spatial information of each of the four test event sequences (i.e., whether a spatial gap was included), the Markov condition, or the novelty of entire test event sequences.

**Procedure.** The procedure for Experiment 3 was identical to that for Experiments 1 and 2; that is, adults were shown the left-to-right training event three times and the right-to-left training event three times. Following the training phase, subjects were shown each test event once in counterbalanced order using a Latin square and asked to rate how perceptually and causally similar each test event was to the training sequence.

## Results

The first analysis examined whether subjects' ratings of the test events differed based on whether they were asked to provide causal or perceptual ratings first (order: perceptual v. causal) and on whether there was an effect of question type, per se (question type: perceptual v. causal). A linear mixed-effects model indicated neither a main effect of order, *F*(1, 222) = .92, *p* = .34, or question type, *F*(1, 30) = .55, *p* = .47, nor an interaction between order and question type, *F*(1, 222) = .04, *p* = .83. This analysis indicates that adults' causal and perceptual ratings of the four test events could not be attributed to order effects or to the type of question they were asked. Like Experiments 1 and 2, subjects' perceptual and causal ratings of the four test events were analyzed separately.

**Primary Analysis**

**Perceptual ratings.** The aim of the first primary analysis was to examine whether adults' perceptual ratings of the four test events differed in their perceptual relationship to the training event. A linear mixed-effects model, with subjects as the random-effects factor and test trial as the within-subjects fixed-effects factor, revealed a significant main effect of test trial, R2= .62, *F*(3, 93) = 33.42, *p* < .0001. Follow-up planned comparisons using permutation testing indicated that subjects provided higher inconsistency ratings of the GFCN test event (*M* = 20.47, Bootstrapped 95% CI[14.59, 26.34]), the GNCF (*M* = 38.81, Bootstrapped 95% CI[30.62, 47]), and GNCN (M = 33.94, Bootstrapped 95% CI[26.55, 41.32]) than to the GFCF test event (*M* = 2.81, Bootstrapped 95% CI[.45, 5.18]), all *p*'s < .001. That subjects' perceptual inconsistency ratings of the GFCF were lower than their ratings of the GFCN, GNCF, and GNCN test events is consistent with that the view that the GFCF was perceived as perceptually familiar with the training sequence. In this manner, subjects’ ratings of the GFCF test event served as a manipulation check to ensure that subjects recognized that this test event was identical to the training sequence.

Likewise, subjects provided higher inconsistency ratings of the GNCF and the GNCN test events than the GFCN test events, all *p*'s < .025. In contrast, subjects’ inconsistency ratings of the GNCF and GNCN test events did not differ reliably, *p* = .19. That subjects' inconsistency ratings for the GNCF and the GNCN test events were higher than those for the GFCF and the GFCN test events suggests that subjects attended to the perceptual features of the training events rather than to the causality of it. In other words, subjects perceived as more inconsistent the test events that included a spatial gap (i.e., GNCF, GNCN) regardless of whether the causal role of the green object in any of the test sequences was consistent with the training sequence. This result is consistent with that in Experiment 2 in which adults provided higher inconsistency ratings to the test events that incorporated a spatial gap.

**Causal ratings.** The aim of the second analysis was to examine whether subjects' causal ratings of the four test events differed. A mixed-effects linear model, with subjects as the random-effects factor and test trial as the within-subjects fixed-effects factor, revealed a significant main effect of test trial, R2 = .57, *F*(3, 93) = 24.47, *p* < .0001. Follow-up planned comparisons indicated that subjects provided higher inconsistency ratings of the GFCN (*M* = 19.06, Bootstrapped 95% CI[12.48, 25.64]), GNCF (*M* = 33.66, Bootstrapped 95% CI[26.39, 40.92]), and GNCN (*M* = 29.63, Bootstrapped 95% CI[23.32, 35.93]) test events than the GFCF test event, *p* < .01. Consistent with the analysis for the perceptual question, this analysis indicates that the GFCF test event was perceived as causally familiar to the training sequence. Likewise, subjects provided reliably higher inconsistency ratings of the GNCF and the GNCN test events than the GFCN test event, both *p*'s < .05. In contrast, subjects’ inconsistency ratings of the GNCF and GNCN test events did not differ reliably, *p* = .22. Similar to subjects' perceptual ratings, the test events that included a spatial gap (i.e., GNCF, GNCN) were perceived as more inconsistent than the test events that did not include a spatial gap, regardless whether the causal role of the green object changed between training and test. Figure 10 shows subjects’ inconsistency ratings of the four test events for the causal and perceptual question.

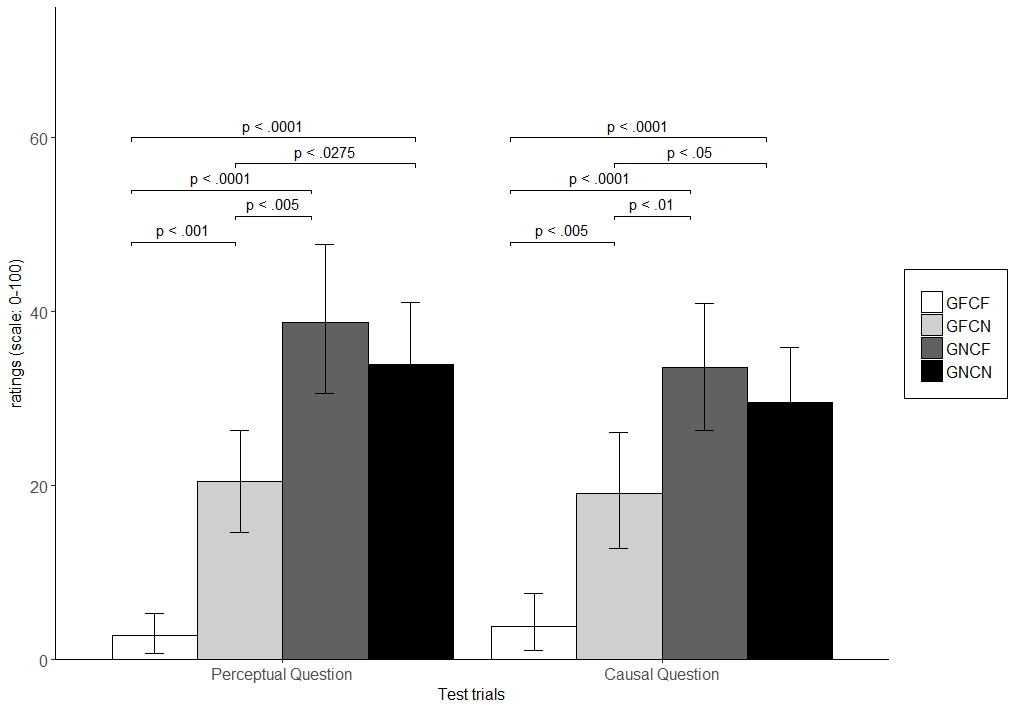


Fig. 10. Subjects’ mean perceptual and causal ratings of the four test events in Experiment 3. The black bars correspond to subjects’ perceptual ratings of the test events and the gray bars correspond to subjects’ causal ratings of the test events.

**Individual Differences**

Figure 11 shows the inconsistency ratings of the four test events for the perceptual question and the causal question for each of the 32 total subjects in the experiment. An analysis of individual differences for the perceptual question revealed that a larger proportion of subjects were classified as Perceptual processors (N = 23; 71.88%) than as "Markov" processors (N = 2; 6.25%) or as Other processors (N = 7; 21.88%), binomial tests, both *p*’s < .01. Of the 23 subjects who were classified as Perceptual processors, 19 processed the events in terms of spatial contiguity information; that is, subjects were sensitive to the fact that the objects in the training sequence only moved upon contact from other objects (i.e., were spatially contiguous) and deemed the test events that incorporated a spatial gap to be inconsistent with the training sequence. In contrast, the proportion of subjects who were classified as Other processors did not differ reliably from the proportion of subjects who were classified as Markov processors, binomial test, *p* = .18.

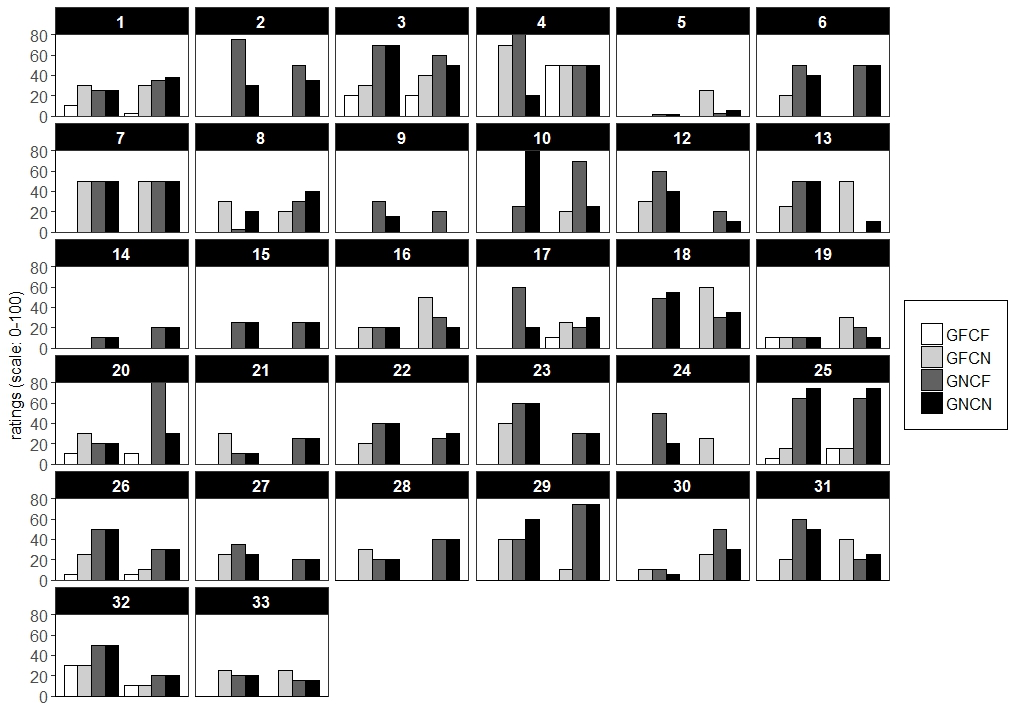


Fig. 11. Each of the 32 subjects’ causal and perceptual ratings of the four test events in Experiment 3.

Likewise, for the causal question, a larger proportion of subjects were classified as Perceptual processors (N = 26; 81.25%) than as Markov processors (N = 2; 6.25%) or Other processors (N = 4; 12.5%), binomial tests, both *p*’s < .0001. Of the 26 subjects who were classified as Perceptual processors, 20 processed the events in terms of spatial contiguity information. In contrast, the proportion of subjects who were classified as Other processors did not differ reliably from the proportion of subjects who were classified as Markov processors, binomial test, *p* = .18.

## Discussion

The purpose of Experiment 3 was to examine whether adults could encode the Markov condition in a novel context that consisted of a single salient object and to determine on what basis adults assigned higher inconsistency ratings to test events that incorporated a spatial gap in an effort to clarify the results of Experiment 2. The results showed that adults provided higher inconsistency ratings to the test events that incorporated a spatial gap than to the test events in which the causal role of the green salient object changed relative to the training phase. This result is consistent with the results from Experiment 2 in that it indicates that adults processed the four-object launching events in terms of low-level perceptual features rather than in terms of causality or the Markov condition or Markov-equivalence classes. This finding also suggests that the reason adults assigned higher inconsistency ratings to the test events that included a causal gap in Experiment 2 may have been because these test events were perceptually rather than causally distinct from the initial training events.

The results from Experiment 1, 2, and 3 together indicate that adults process complex causal events such as those used here in terms of low-level perceptual information rather than in terms of the Markov condition. The finding that adults did not encode the Markov condition or show sensitivity to Markov-equivalence classes in Experiments 2 and 3 and only under limited conditions in Experiment 1 suggests that their ability to represent causal events as CGMs is restricted and may depend on the perceptual features in the event. One potential criticism of these results, especially those from Experiments 2 and 3, is that the reason adults showed little sensitivity to the Markov condition and Markov-equivalence classes was because the perceptual manipulations to the test sequences in these experiments were so great that they disrupted adults' sensitivity to these cues. Although we discuss this criticism in some detail in the General Discussion, it is important to note here that proponents of CGMs maintain that adults’ sensitivity to the Markov condition and Markov-equivalence classes should be robust to salient perceptual manipulations so long as such manipulations do not affect the set of underlying conditional independencies (e.g., Gopnik & Tenenbaum, 2007).

Given that Experiments 1, 2, and 3 has consistently shown that a small proportion of subjects showed sensitivity to the Markov condition and Markov-equivalence classes, one unresolved issue is that it is unclear at what level adults encode the Markov condition when shown four-object launching event sequences. In other words, it is unknown from these results over what dimensions adults encoded the Markov condition and formed Markov-equivalence classes. For example, it is possible that adults encoded the Markov condition and attended to Markov-equivalence classes over the shape dimension, color dimension, both, or neither. Likewise, it is unclear whether, in the presence of competing dimensions adults will selectively encode the Markov condition over some dimensions but not over others. We explored this issue in Experiment 4.

# Experiment 4

Experiment 4 was designed to examine whether, and over what dimensions, adults encode the Markov condition and whether they show sensitivity to the Markov condition and Markov-equivalence classes when asked to make causal and perceptual comparisons. Experiment 4 differed from Experiments 1-3 in one crucial way; that is, unlike Experiments 1-3 that used objects that were the same shape and sometimes the same color, Experiment 4 used four objects that differed both in terms of their shape as well as in terms of their color. Given that the aim of Experiment 4 was to determine whether sensitivity to the Markov condition and Markov-equivalence classes privileges particular perceptual dimensions, it was necessary to include test events in which it was possible independently to manipulate the Markov condition defined over the shape or color dimension. Thus, if sensitivity to the Markov condition in the training events privileges the color dimension, then test events that incorporate the same conditional independencies defined over the color dimension should be perceived as more familiar than test events that incorporate different conditional independencies defined over the same dimension. In contrast, if sensitivity to the Markov condition in the training events privileges the shape dimension, then test events that incorporate the same conditional independencies defined over that dimension should be perceived as more familiar than events that incorporate different conditional independencies defined over the same dimension.

## Methods

**Subjects.** Thirty-two college students were recruited from Carnegie Mellon University to participate in Experiment 4. Data about age or sex of participants were not collected.

**Stimuli and Design.** Similar to the first three experiments, the stimuli in Experiment 4 were computer-animated launching-event sequences and the stage on which the events were depicted was the same as that in Experiments 1-3. The dimensions of the shapes used in this experiment was also identical to those used in Experiments 1-3, although the actual shape of each object differed across objects (discussed below).

During training, subjects were shown a four object launching sequence (Fig. 12) in which a brown hexagon—which was initially out of sight and off screen—moved horizontally across the stage until it made contact with a blue pentagon that was located a short distance away. On contact, the brown hexagon stopped moving and the blue pentagon began horizontally to move until it contacted a red upside down triangle. Following contact from the blue pentagon, the upside down triangle began horizontally to move until it made contact with and subsequently caused a final green circle horizontally to move until it was off stage and out of sight. The entire training event sequence lasted approximately 10 s and each sequence was played both from left to right and from right to left to eliminate potential direction effects.

In the test phase, subjects were shown four test events: SFCF (Shape-Familiar; Color-Familiar), SNCF (Shape-Novel; Color-Familiar), SFCN (Shape-Familiar; Color-Novel), and SNCN (Shape-Novel; Color-Novel). In the SFCF test event, the ordering of the shapes and colors (from left to right) was consistent (i.e.,familiar) with the training sequences. In the SNCF test event, the ordering of the shape stimuli differed from that shown during training, whereas the ordering of the colors was consistent with that shown during training. In the SFCN test event, the ordering of the colors differed from that shown during training, whereas the ordering of the shapes was consistent with that shown during training. Finally, in the SNCN test event both the ordering of the colors and shapes differed from that shown during initial training. These test events are depicted in Figure 12.

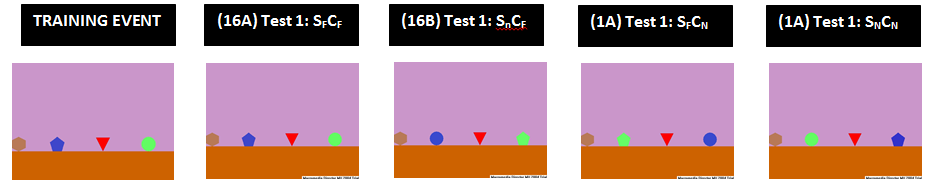


Fig. 12. Example of the training event and test events used in Experiment 4.

Critically, the order in which the shapes or colors appeared and moved corresponded to whether the Markov condition for this particular dimension was preserved from training to test. Thus, if the order of the shapes differed in the test phases from what was observed during training, then the Markov condition defined over the shape dimension violated that shown during initial training. In contrast, if the order of the colors differed in the test phases from what was observed during training, then the Markov condition defined over the color dimension violated that shown during the initial training phase. Finally, if both the order of shapes and colors differed in the test phases from what was observed during the initial training phase, then the Markov condition defined over the shape and color dimensions violated that shown during the initial training phase.

## Predictions

It was predicted that if adults selectively encoded the Markov condition defined over the shape dimension and ignored the Markov structure defined over the color dimension, then they should assign higher inconsistency ratings to the test events in which the ordering of the shapes violated the ordering observed during the training phase; that is, adults should assign higher inconsistency ratings to the SNCF and SNCN test events than to the SFCF and SFCN test events. In contrast, if adults encoded the Markov condition defined over the color dimension and ignored the Markov structure defined over the shape dimension, then they should assign higher inconsistency ratings to the test events in which the ordering of the colors violated the ordering of the colors in the training event, regardless of whether the ordering of the shapes was consistent with that shown during initial training; in other words, adults should assign higher inconsistency ratings to the SFCN SNCN test events than to the SFCF and SNCF test events. However, if adults failed to encode the Markov condition over either dimension and instead responded on the basis of perceptual novelty, then they should assign higher inconsistency ratings to the SFCN, SNCF, and SNCN test events than to the SFCF test event. Finally, if subjects represented the training events as launching events in the most general sense, then their inconsistency ratings of all four test events should be at floor because all four test events are instances of a more general launching-event sequence. These predictions are illustrated in Figure 13.

Fig. 13. Predicted pattern of responding to the test events in Experiment 4 based on whether subjects responded to the test events based differences relative to the training phase in the causal roles of the four objects, color of the four objects, novelty of an entire test event sequences, or familiarity of an entire test event sequences.

**Procedure**

The procedure for Experiment 4 was identical to that for Experiments 1-3; that is, adults were shown the left-to-right training event three times and the right-to-left training event three times. Following the training phase, subjects were shown each test event once in counterbalanced order using a Latin square and were asked to rate how perceptually and causally similar each test event was to the training sequence.

## Results

The first analysis examined whether subjects' ratings of the test events differed based on whether they were asked to provide causal or perceptual ratings first (order: perceptual v. causal) and on whether there was an effect of question type, per se (question type: perceptual v. causal). A linear mixed-effects model revealed no main effect of question type, *F*(1, 222) = .36, *p* = .55, but there was a main effect of order that further was qualified by an interaction between order and test type, *F*(1, 222) = 4.59, *p* < .05. Follow-up comparisons indicated that subjects who were asked the perceptual question first provided higher inconsistency ratings when asked to make perceptual comparisons (*M* = 31.19, Bootstrapped 95% CI[26.6, 35.77]) than when asked to make causal comparisons (*M* = 24.38, Bootstrapped 95% CI[17.68, 31.07]), *p* < .0001. In contrast, the inconsistency ratings for subjects who were asked the causal question first did not differ when asked to make causal comparisons (*M* = 20.78, Bootstrapped 95% CI[16.18, 25.38]) and perceptual comparisons (*M* = 16.95, Bootstrapped 95% CI[11.23, 22.68]), *p* = .34. That adults who were asked the causal question first provided higher inconsistency ratings when asked to make causal comparisons than when asked to make perceptual comparisons (although it was not statistically significant) and adults who were asked the perceptual question first provided higher inconsistency ratings when asked to make perceptual comparisons than when asked to make causal comparisons suggests that adults may have been more confident about their initial ratings than their second ratings. One reason this pattern may have emerged is because the design of Experiment 4 was more complicated than that of Experiments 1-3 given that subjects were asked to reason about events in which the colors and shapes of the objects changed in the test events.

**Primary Analysis**

**Perceptual ratings.** Figure 14 shows subjects’ inconsistency ratings to the four test events for the causal and perceptual questions. The aim of the first primary analysis was to examine whether and to what extent subjects' perceptual inconsistency ratings differed for each of the four test events. A linear mixed-effects model, with subjects as the random-effects factor and test trial as the within-subjects fixed-effects factor, revealed a marginally significant main effect of test trial, R2 = .65, *F*(3, 93) = 2.16, *p* = .098. Given that the main effect was marginally significant and that 65% of the variance was accounted for by the predictors in the model, planned comparisons were undertaken. This analysis indicated that subjects provided reliably higher inconsistency ratings of the SNCF test event (*M* = 27.66, Bootstrapped 95% CI[19.76, 35.55]) than of the SFCF test event (M = 18.375, Bootstrapped 95% CI[10.14, 26.61]), *p* < .05, and marginally higher inconsistency ratings of the SFCN (M = 25.63, Bootstrapped 95% CI[18.19, 33.05]) and SNCN (M = 24.63, Bootstrapped 95% CI[17.42, 31.83]) test events than of the SFCF test event, all *p*'s < .15. Follow-up Bayes’ Factor analyses indicated that the observed difference between the causal ratings of the SFCN and the SFCF test events (∆SFCN -SFCF = 7.25) was 5.16 times more likely under the alternative hypothesis than the null hypothesis. Likewise, the observed difference between the causal ratings of the SNCN and the SFCF test events (∆SNCN -SFCF = 6.25) was 3.58 times more likely under the alternative hypothesis than the null hypothesis Using Raftery (1995) and Jeffrey'(1961) guidelines for interpreting Bayes’ Factors, both Bayes’ Factors represents positive and substantial evidence in support of the alternative hypothesis. In contrast, subjects' inconsistency ratings of the SNCF, SFCN, and SNCN test events did not differ reliably from each other, all *p*'s > .5.

**Causal ratings.** The aim of the second primary analysis was to examine whether and to what extent subjects' causal inconsistency ratings differed for each of the four test events. Similar to the first analysis, a linear mixed-effects model, with subjects as the random-effects factor and test trial as the within-subjects fixed-effects factor, revealed a significant main effect of test trial, R2 = .7, F(3, 93) = 5.85, *p* < .005. Follow-up planned comparisons using permutation testing indicated that subjects provided reliably higher inconsistency ratings of the SNCF (*M* = 23.81, Bootstrapped 95% CI[15.55, 32.07]), SFCN (*M* = 24.59, Bootstrapped 95% CI[16.49, 32.7]), and SNCN (*M* = 28.44, Bootstrapped 95% CI[19.98, 36.89]) test events than of the SFCF (*M* = 13.47, Bootstrapped 95% CI[6.68, 20.26]), all *p*'s < .05. In contrast, subjects' inconsistency ratings of the SNCF, SFCN, and SNCN test events did not differ reliably from each other, all *p*'s > .4.

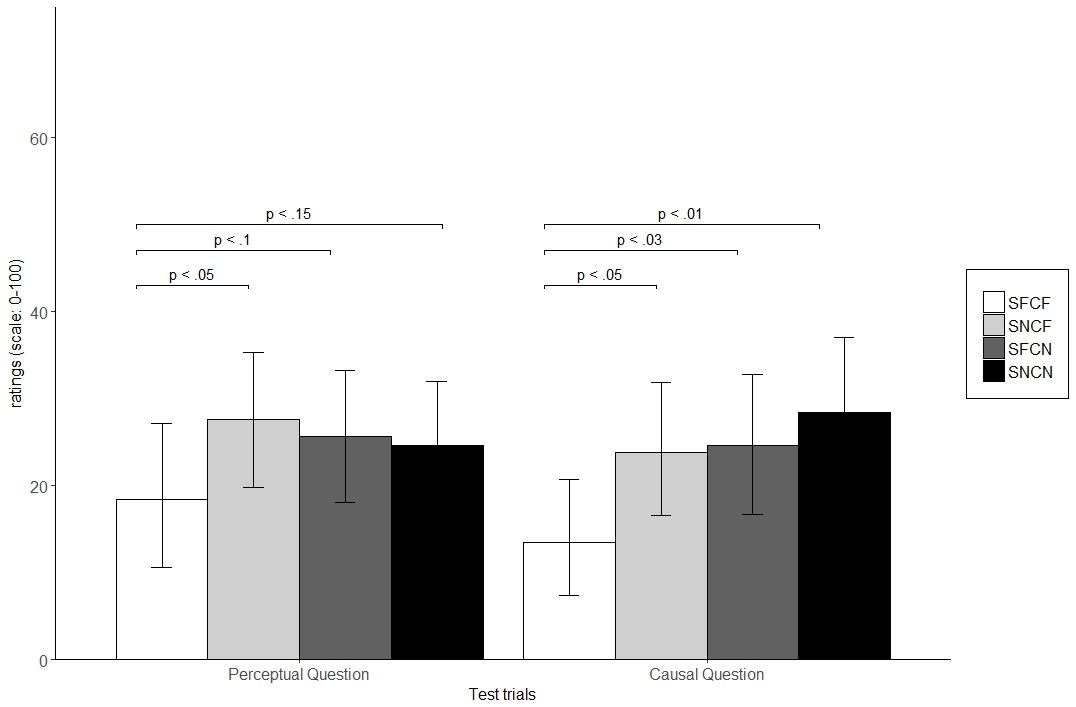


Fig. 14. Subjects’ mean perceptual and causal ratings of the four test events in Experiment 4. The black bars correspond to subjects’ perceptual ratings of the test events and the gray bars correspond to subjects’ causal ratings of the test events.

These two analyses suggest that subjects neither encoded the Markov condition defined over the shape dimension nor over the color dimensions. These results suggest instead that adults responded to the test events based on perceptual novelty. This is based on the fact that adults provided higher inconsistency ratings to the test events that introduced a novel perceptual change relative to the test event that was identical to that shown during the training phase.

**Individual Differences**

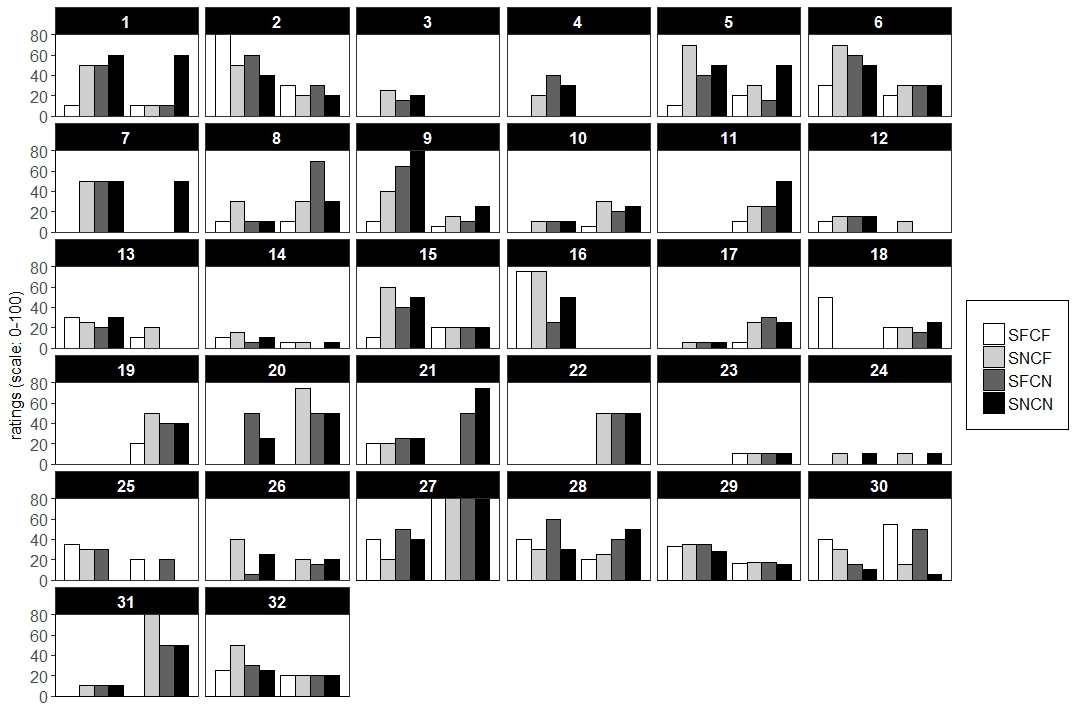


Fig. 15. Each of the 32 subjects’ mean causal and perceptual ratings of the four test events in Experiment 4.

An analysis of individual differences for the perceptual question revealed that a larger proportion of subjects were classified as Perceptual processors (N = 17; 53.13%) than as Other processors (N = 6; 18.75%), binomial test, *p* < .05, and a marginally larger proportion of subjects were classified as Perceptual processors than as Markov processors (N = 9; 18.75%), binomial tests, *p =* .17. However, the proportion of subjects who were classified as Other processors did not differ reliably from the proportion of subjects who were classified as Markov processors, binomial test, *p* = .61. Of the 17 subjects who were classified as Perceptual processors, 13 subjects responded to the test events based on any perceptual novelty; that is, 13 subjects provided higher inconsistency ratings to the last three test events presumably because these test events incorporated at least one novel aspect relative to that shown during initial training.

Likewise, for the causal question, a larger proportion of subjects were classified as Perceptual processors (N = 18; 56.25%) than as Markov processors (N = 2; 6.25%) or Other processors (N = 7; 21.88%), binomial tests, both *p*’s < .05. However, the proportion of subjects who were classified as Other processors differed only marginally from the proportion of subjects who were classified as Markov processors, binomial test, *p* = .18. Of the 18 subjects who were classified as Perceptual processors, 12 subjects processed the events in terms of spatial contiguity information. These analyses, similar to the previous analyses of individual differences, suggest that adults show a greater tendency to process complex causal events that consist of four objects in terms of low-level perceptual and associative cues than in terms of the Markov condition and Markov-equivalence classes.

## Discussion

The goal of Experiment 4 was to determine whether and over what dimensions adults can encode the Markov condition. The results revealed that adults did not encode the Markov condition defined over the color dimension or the shape dimension. Rather, the results revealed that adults responded to the test events on the basis of perceptual novelty. This overall finding was further corroborated in an individual difference analysis that revealed that a larger proportion of adults responded based on perceptual novelty than based on familiarity or the Markov condition defined over the color or shape dimension for both the perceptual and causal questions. Similar to the results from Experiments 1-3, these results suggest (a) that adults process complex causal events that consist of four objects in terms of perceptual features and cues, (b) that their representations of such events may be grounded in, and built from, low-level perceptual cues that are readily available in the input, and (c) that sensitivity to the Markov condition and Markov-equivalence classes is limited to particular contexts such as that in Experiment 1 when subjects were asked to make causal comparisons.

One unresolved is when sensitivity to the Markov condition and Markov equivalence classes emerges. This is an important question to answer because it has the potential to address whether such sensitivity is present from late infancy—and thus does not change between late infancy and adulthood—or whether such sensitivity develops perhaps as a function of experience with causal events in the real world. Thus, Experiment 5 was designed to examine whether infants between 18 and 22 months of age show sensitivity to the Markov condition and Markov-equivalence classes, and whether infants represent causal events as CGMs or in terms of low-level perceptual features and cues. Although the ability to perceive causality in launching event sequences emerges by 6½ months of age (e.g., Leslie & Keeble, 1987), arguably it is not until much later in development that sensitivity to the Markov condition and Markov equivalence classes emerges. This is because—unlike causal perception that requires infants to distinguish between agents and recipients and causal and non-causal events—sensitivity to the Markov condition and Markov equivalence classes requires that infants attend not only to causality but to conditional probability information that relate the objects within such sequences, which presumably emerges by the time that infants can reason about causal events. Thus, Experiment 5 tested infants between 18 and 22 months of age because it is not until infants are around 18 months of age that they can reason about causal events and presumably the conditional probability information contained within them (e.g., Sobel & Kirkham, 2006). Eighteen- to 22-month-olds were also tested because it is not until around 15 months of age that infants can perceive causality in causal-chain launching-event sequences that are composed of three objects.

# Experiment 5

The aim of Experiment 5 was identical to Experiment 1; that is, it was designed to examine on what basis learners process four-object launching-event sequences, whether learners are sensitive to and can encode the Markov condition, and whether they can form Markov-equivalence classes. However, Experiment 5 differed from Experiments 1 to 4 in that it examined at what age sensitivity to the Markov condition and Markov equivalence classes in complex causal sequences emerges—and thus at what age the ability to represent complex causal events as CGMS emerges—and whether such sensitivity undergoes developmental change between late infancy and adulthood. Experiment 5 also differed from Experiments 1 to 4 in that it used a habituation procedure to assess infants’ sensitivity to the Markov condition and Markov-equivalence classes. The same predictions that were discussed in Experiment 1 applied to Experiment 5, except that looking time was used as the dependent measure rather than subjects’ inconsistency ratings. In terms of predicted looking times, infants were expected to look longer at the GBGR, GBRG, and GRBG test events if they were sensitive to the Markov condition and Markov equivalence classes, whereas they were expected to look longer at a different subset of test events if they processed the test events in terms of low-level perceptual features or in terms of other cues.

## Method

**Subjects.** The subjects were 25 infants at 18 months of age (mean age 18 months 8 days; range = 12 months to 33 months 29 days). Thirteen males and 12 females participated in this experiment. Data from an additional 9 infants were excluded from the final sample, 4 due to fussing or crying, 2 to technical problems, and 3 to experimenter error. Eleven infants were randomly assigned to blue group and 14 infants were randomly assigned to the red group; that is, 11 infants were habituated to the BGRG event, and 14 infants were habituated to the RGBG event. Infants were recruited through birth lists obtained from a private company and were given a small gift for their participation.

**Stimuli and Design.** The habituation and test stimuli were the same as those in Experiment 1. In particular, the BGRG and RGBG events were used as the habituation stimuli, and the GBGR, GRGB, GBRG, GRBG events were used as test stimuli. The length of time it took the habituation and test events to be completed were approximately 8 s, and each event could be repeated up to three times per habituation trial. Pretest and posttest trials were also included to distinguish between infants who showed a lack of alertness to the test trials and those who showed a lack of learning throughout the experiment. In both events, a wind-up toy mouse moved horizontally across the screen until it was off-screen. This event lasted for 7½ seconds. Individual presentations of each event were separated by a blue screen that descended and ascended over a period of 2 s. Finally, an expanding and contracting green circle—which was in the center of a black screen—was played to redirect infants to the monitor whenever they looked away. The attention-getting stimulus was played at the beginning of the experiment and after each habituation and test trial. A bell sound was presented in synchrony with this movement to maximize the attractiveness of the event and to secure the infant's visual attention.

**Apparatus.** Each infant was tested individually in a small, silent, dimly lit laboratory room (~3.0 x 2.5 m). During the experiment, events appeared on a 14-inch by 24-inch LCD monitor, which was placed approximately 48 inches away from the infant's face. The monitor was situated on a table and was in plain sight to the infant and their caregiver. Two speakers were placed on either side of the monitor to play the attention-getting sound. Surrounding the test chamber and at the rear of the monitor was a black curtain that spread from the ceiling to the floor and across the infants' visual field from left to right. A closed-circuit video camera was used to record infants' responses to the habituation and test stimuli. This camera was placed out of view and above the monitor to record infant's looking behavior. The experiment was controlled by the HabitX 2000 software program (Cohen, Atkinson, & Chaput, 2000) on an Apple computer.

An experimenter—who was out of sight and behind the curtain—observed each infant's visual gaze on a second monitor. This second monitor allowed the experimenter to record infants' looking behavior. Experimenters were trained to press and hold a particular key on the computer keyboard whenever infants oriented to the computer monitor. Experimenters were instructed to release this key whenever the infant looked away from the monitor. Upon fixating the computer monitor, the experimenter began the next trial by pressing a preset key on the computer keyboard. The computer recorded the length of each key press, and thereby the amount of time infants’ fixated a particular habituation or test event. Experimenters were blind to the exact event presented to each infant given that it was possible for each infants habituated at different rates.

**Procedure.** Infants were tested one at a time in a quiet room and each infant sat on his or her caregiver's lap and faced a monitor on which the stimuli were displayed. Each caregiver was instructed to remain neutral, to abstain from interacting with the infant, and to wear opaque sunglasses throughout the course of the experiment to ensure that infants' responses to the stimuli were natural and unperturbed. Each caregiver was blind to the experimental hypotheses and predictions and were told only that this experiment was designed to examine how infants process complex causal-event sequences. This means that it was unlikely that the caregiver could reliably influence an infants’ behavior during experiment.

An infant-controlled habituation procedure was used (Cohen, 1973). The experiment consisted of a pretest trial, a maximum of 16 habituation trials, four test trials, and one posttest trial. Trials terminated when an infant looked away from the monitor for one second or had accumulated 32 seconds of continuous looking.

During the habituation phase of the experiment, infants were shown the BGRG and RGBG events, which were presented in semi-random order, with half of the infants receiving the BGRG event first and the other half receiving the RGBG event first. An infant habituated when their looking time decreased to a set criterion level or until 16 trials had been presented; that is, infants habituated when their looking times on a block of three successive trials was less than 50% of the total looking time on the first three trials. Following habituation, the test phase began. All infants reached this criterion and, as such, were included in all subsequent analyses. The order of the four test events—that is, GBGR, GRGB, GBRG, GRBG—was counterbalanced using a Latin square.

## Results

Given that there was evidence of non-normality in infants' looking times to the test events, non-parametric analyses (with 4,000) replications were conducted for hypothesis testing and to estimate confidence intervals. Because preliminary analyses indicated that there was no main effect of sex of the infant, order of the test trials, group type (i.e., red or blue group), or any interaction between the three factors on infants’ looking times to the test events, the data were collapsed across the levels of these factors. Note that this analysis indicates that differences in test trial looking cannot be attributed to sex, order of test trials, and group type.

The first analysis compared infants' looking times to the pretest and posttest event as a function of sex, order of the test trials, and group type. The rationale behind this analysis was to determine whether, in the case of equal looking times to the test events, such looking behavior resulted from general fatigue or a lack of learning. It is generally believed that a significant decrease in looking times to the postest event relative to those in the pretest trials, in combination with low looking times to the four test events, suggests that infants became fatigued in the experiment (Werker et al., 1998). If infants' looking to the postest event is elevated relative to the four test events used here, this would indicate either that (a) infants failed to process the test events or (b) processed the events as equivalent in an unspecified way. Thus, infants' looking times were analyzed with a mixed-effects four-way ANCOVA with trials (pretest vs. posttest) as the within-subjects factor and sex (male vs. female), order of the test trials, and group type as the between-subjects factors. Age was included as the covariate and subject ID was included as the random-effects factor. The analysis revealed no main effect of sex *F*(1,34) = 0.21, *p* = 0.65, *ηp2* = 0.006, no main effect of test trial order *F*(1,34) = 2.37, *p* = 0.13, *ηp2* = 0.07, no main effect of trials *F*(1,34) = 2.17, *p* = 0.15, *ηp2* = 0.06, but a significant main effect of group type *F*(1,34) = 4.91, *p* = .03, *ηp2* = 0.13. There was no significant interaction between any of these variables, all *p*'s > 0.08, all *ηp2*'s > 0.08.

To determine whether the reliable main effect of group type was practically meaningful, a follow-up Bayes' Factor was estimated. This analysis indicated that the data were 1.77 times more likely under the alternative hypothesis that looking times to the pretest and postest events differed between the red and blue groups than under the null hypothesis that there was no difference in looking times to the events. Using Raftery (1995) and Jeffrey'(1961) guidelines for interpreting Bayes’ Factors, a Bayes’ Factor of 1.77 represents weak and anecdotal evidence in support of the alternative hypothesis. Thus, the differences in looking times to the pretest and postest events—although statistically significant—was not reliably significant or meaningful. Importantly, that the main effect of trials was not significant indicates that infants' looking times to the pretest and postest events did not differ.

**Habituation Phase.** On average, infants required 6.24 trials (Bootstrapped 95% CI[5.81, 6.67]) to habituate and accumulated 81.35 seconds (Bootstrapped 95% CI[72.59, 90.12]) of total looking time during habituation. To examine the effects of sex, group, and test trial order on total looking to the habituation trials, a 2(sex: male, female) x 2(group: blue, red) x 2(test trial order: GBGR/GRGB/GBRG/GRBG vs GRGB/ GRBG/GBGR/GBRG,) was conducted with subject as the random-effects factor. The results indicated no main effect of sex *F*(1,17) = 0.15, *p* = 0.69, *ηp2* = 0.009, no main effect of group *F*(1,17) = 0.02, *p* = 0.88, *ηp2* = 0.001, no main effect of order of test trials *F*(1,17) = 0.3, *p* = 0.59, *ηp2* = 0.02, and no interaction between any of the three factors, all interaction *p*'s > 0.33, all interaction *ηp2*’s < .06.

Likewise, to examine the effects of sex, group, and test trial order on number of trials to reach the habituation criterion, a 2(sex: male, female) x 2(group: blue, red) x 2(test trial order: 1234, 2413) was conducted with subject as the random-effects factor. This analysis also did not yield a main effect of sex *F*(1,17) = 1.31, *p* = 0.27, *ηp2* = 0.07, no main effect of group *F*(1,17) = 0.15, *p* = 0.71, *ηp2* = 0.009, no main effect of order of test trials *F*(1,17) = 1.65, *p* = 0.22, *ηp2* = 0.09, and no interaction between any of the three factors, all interaction *p*'s > 0.54, all interaction *ηp2*’s < .01.

The final two analyses were similar to the first two analyses except that age was included as a covariate and subject ID as a random-effect in a mixed-effects ANCOVA model. The first analysis—which examined whether there was an effect of sex, group type, or test trial order on infants' amount of total looking to the habituation events when age was controlled for—did not yield a significant main effect of sex *F*(1,17) = 0.04, *p* = 0.84, *ηp2* = 0.002, main effect of group *F*(1,17) = 0.0004, *p* = 0.98, *ηp2* = 0.00003, main effect of test trial order *F*(1,17) = 0.39, *p* = 0.54, *ηp2* = 0.02, and no interaction between any of the three factors, all interaction *p*'s > 0.25, all interaction *ηp2* < 0.08.

Likewise, the second analysis—which examined whether there was an effect of the same three categorical predictors on the number of trials to reach the habituation criterion when age was included as a covariate and subject ID as a random effect—did not yield a significant main effect of sex *F*(1,17) = 0.96, *p* = .34, *ηp2* = 0.05, main effect of group *F*(1,17) = 0.6, *p* = 0.81, *ηp2* = .004, main effect of test trial order *F*(1,17) = 2.09, *ηp2* = 0.11, and no interaction between any of the three factors, all interaction *p*'s > 0.34, all interaction *ηp2*’s < 0.05.

These habituation analyses indicate that the number of trials to reach the habituation criterion and amount of total looking time to the habituation events did not depend on the infants' sex, the order of the test trials or on whether they were shown events in which the red (or blue) circle was the first object in the habituation sequences. Thus, any looking time differences to the test events cannot be attributed to differences in infants’ responses to the habituation events for any of the three factors. In other words, the absence of any main effect or interaction indicates that habituation patterns were uniform across the three factors.

**Test Phase.** Infants' mean looking times to the four test events as well as to the posttest event are illustrated below in Figure 16. Because Experiment 5 tested infants between the ages of 18 and 22 months of age, age was controlled for by including it as a covariate in the main analysis. Thus, a mixed-effects ANCOVA was conducted with trials (GBGR, GRGB, GBRG, GRBG, posttest) as the within-subjects factor, age as the covariate, and subject ID as the random-effects factor. This analysis revealed significant main effect of trials after controlling for age *F*(4,96) = 12.64, *p* < .0001, *ηp2* = .25. Planned comparisons using permutation tests showed that infants looked equally long at the GBGR (*M* = 9.53, Bootstrapped 95% CI[6.52, 12.55]), GRGB (*M* = 7.42, Bootstrapped 95% CI[4.64, 10.21]), GBRG (*M* = 8.96, Bootstrapped 95% CI[5.48, 12.44]), and GRBG (*M* = 7.22, Bootstrapped 95% CI[4.09, 10.34]). However, they looked significantly longer at the posttest event (*M* = 20.28, Bootstrapped 95% CI[16.17, 24.39]) than at the GBGR, GRGB, GBRG, and GRBG test events, all *p*’s < 0.0001. That infants looked equally long at all four test events and reliably longer at the posttest event than at any of the test events suggests that their failure to process the test events in terms of the Markov condition or low-level features did not result from a lack of alertness and general fatigue. Instead, these results suggest that infants perhaps may have processed the test events as “a chain of objects that collide with and cause other objects to move.”

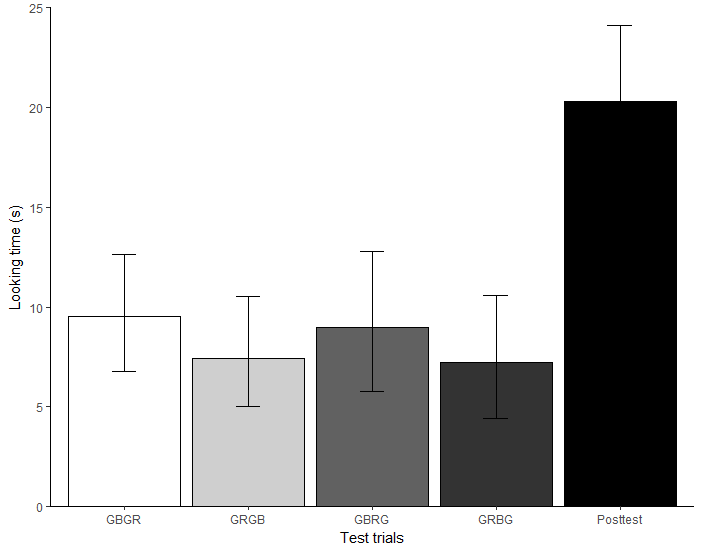


Fig. 16. Infants’ mean looking times to the four test events and one posttest event in Experiment 5.

**Individual Differences.** The aim of the next several analyses was to examine to what extent age and habituation rate—both defined according to a median split—influenced infants’ looking behavior to the four test events. The first analysis was undertaken because sensitivity to the Markov condition and Markov-equivalence classes appeared to be emerging for older infants, but not for younger, infants when the effect of test trial type on infants’ looking behavior was plotted separately for the younger (<=18.06) and older infants (>18.06) in Figure 17. This conclusion is based on the fact that the distribution of looking times to the four test events for the older infants, but not the younger infants, is qualitatively consistent with the predicted looking times of infants who are sensitive to the Markov condition and Markov-equivalence classes.

The second analysis was undertaken because previous research has indicated that it is sometimes the case that infants who require fewer trials to habituate tend to perform better on certain cognitive tasks than do infants who require more trials (e.g., Baillargeon, 1987); that is, it is conceivable that "fast habituators,” but not "slow habituators," in the age range tested showed sensitivity to the Markov condition and Markov-equivalence classes.

Finally, both analyses were undertaken to determine whether, like adults in Experimenters 1-4, differences existed in the proportion of younger and older infants and fast and slow habituators who processed the events in terms of the Markov condition, low-level perceptual features, or other cues (whose predicted looking-time distributions were inconsistent with that if infants processed the events perceptually or in terms of the Markov condition). The median-split age analysis is reported first and is followed by the habituation-rate analysis.

Note that infants' looking times to the posttest event were not included in any of the subsequent individual-difference analyses because the analysis above indicated that infants were fully alert during the experiment despite the fact that their looking times to the four test events were equivalent. These looking times were also not included because the aim of this section was to determine whether the individual-difference variables of age and habituation rate influenced infants' looking behavior to the four test events but not to the posttest event.

**Median-split Age.** Although the main analysis reported above did not yield a main effect of test trial order when age was included as a covariate, a median split was used nonetheless to examine the effect of age on infants' looking behavior to the four test events; that is, to determine whether the older (or younger) infants were more likely to process the events in terms of the Markov condition. The mean looking times to the four test events, as a function of median-split age, is shown below in Figure 17.

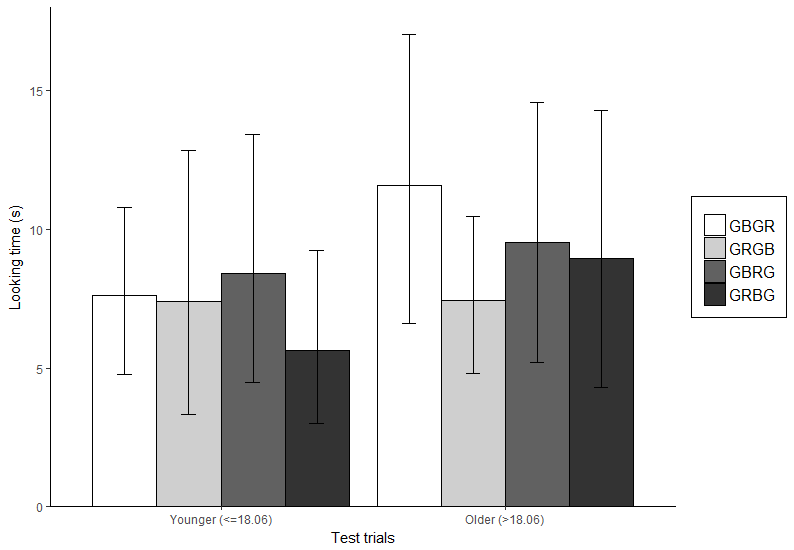


Fig. 17. Mean looking times to the four test events for the “younger” and “older” infants.

A 4(test trial: GBGR, GRGB, GBRG, GRBG) x 2(age: Younger [<= 18.06] vs Older [>18.06]) mixed-effects repeated measure ANOVA was conducted in which test trial was included as a within-subjects factor, median-split age was included as a between-subjects factor, and subject ID was included as the random-effects factor. The results indicated that there was no main effect of age *F*(1,23) = 1.02, *p* = 0.32, *ηp2* = 0.02, no main effect of test trial order *F*(3,69) = 0.64, *p* = 0.59, *ηp2* = 0.02, and no interaction between the two categorical predictors, *F*(3,69) = 0.42, *p* = 0.74, *ηp2* = 0.01. Thus, the older and younger infants looked equally long at the test events. Despite these null results, it is possible that differences existed in the proportion of younger and older infants who attended to the Markov condition, low-level perceptual features, or other cues. This issue was examined separately for the younger and older infants.

**Younger Infants.** Figure 18 shows the looking times for the younger infants (N = 13).An analysis of individual differences indicated that the proportion of younger infants who were classified as Markov processors did not differ reliably from that for infants classified as Perceptual or Other processors, χ2 = 2.92, *p* = 0.27. This analysis indicates that there were no reliable individual differences among "younger" infants in the extent to which they processed the events in terms of low-level perceptual cues, the Markov condition, or other features and cues.

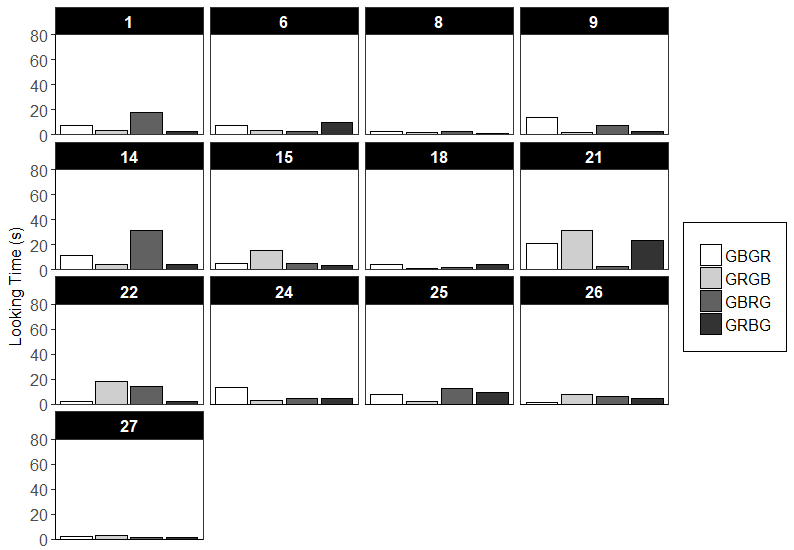


Fig. 18. Mean looking times to the four test events for the 13 “younger” infants.

**Older Infants.** Figure 19 shows the looking times for the older infants (N = 12).Similar to the analysis presented above with younger infants, an analysis of individual differences revealed that the proportion of older infants who were classified as Markov processors did not differ reliably from that for infants classified as Perceptual or Other processors, χ2 = 0.5, p = 0.94. Thus, similar to the analysis reported above with younger infants, this analysis indicates that there were no reliable individual differences in the extent to which older infants were classified as Markov processors, Perceptual processors, or Other processors. In addition, there was no indication that a larger percentage of the older infants showed sensitivity to the Markov condition and Markov equivalence classes than to other cues and features.

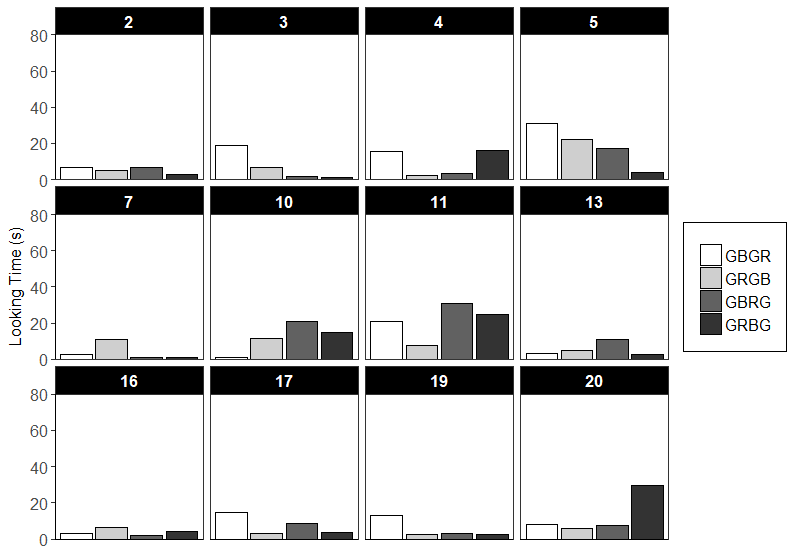


Fig. 19. Looking times to the four test events for the 12 “older” infants.

**Habituation Rate.** To examine the effect of habituation rate on infants' looking behavior to the four test events, a 4(test trial: GBGR, GRGB, GBRG, GRBG) x 2(Habituation rate: fast [<=5 trials] vs slow habituators [>5 trials]) mixed-effects repeated-measures ANOVA was conducted. Test trial was included as a within-subjects factor, median-split habituation rate was included as a between-subjects factor, and subject ID was included as the random-effects factor. The results indicated that there was no main effect of habituation rate *F*(1,23) = 0.003, *p* = 0.96, *ηp2* = 0.00005, no main effect of test trial *F*(3,69) = 0.64, *p* = 0.59, *ηp2* = 0.016, and no interaction between the two factors *F*(3,69) = 0.28, *p* = 0.84, *ηp2* = 0.007. Despite the fact that the fast and slow habituators did not differ in their looking behavior to the four test events, it is possible that differences existed in the proportion of fast and slow habituators who attended to the Markov condition, low-level perceptual features, or other cues. The mean looking times to the four test events, as a function of median-split habituation rate, is shown below in Figure 20.

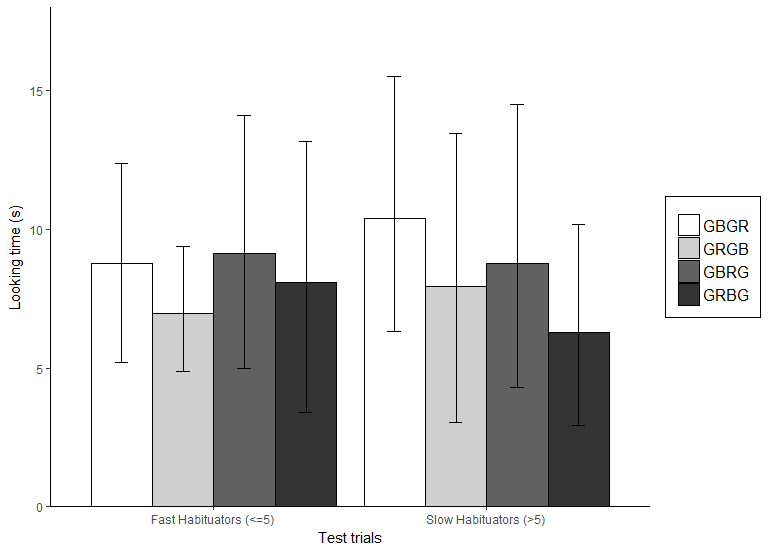


Fig. 20. Looking times to the four test events for the “fast” and “slow” habituators.

**Slow Habituators.** Figure 21 shows the looking times for the slow habituators (N = 12). An analysis of individual differences indicated that the proportion of slow habituators who were classified as Markov processors did not differ reliably from that for slow habituators who were classified as Perceptual or Other processors, χ2 = 2, *p* = 0.42. This analysis indicates that the slow habituators were as likely to process the events in terms of the Markov condition and Markov-equivalence classes as they were to process the events in terms of low-level perceptual features and other cues.

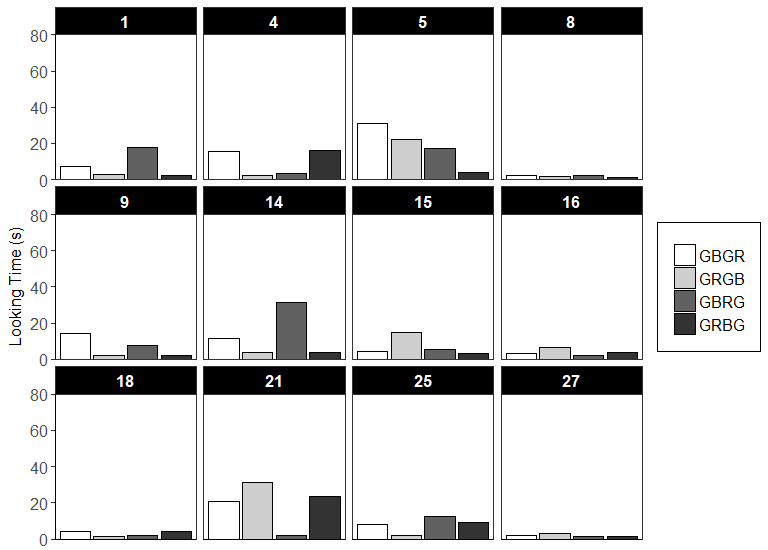


Fig. 21. Looking times to the four test events for the 12 “slow” habituators infants.

**Fast Habituators.** Figure 22 shows the looking times for the fast infants (N = 12). Similarly, an analysis of individual differences indicated that the proportion of fast habituators who were classified as Markov processors did not differ reliably from that for fast habituators who were classified as Perceptual or Other processors, χ2 = 2, *p* = 0.48. This analysis indicates that the fast habituators were as likely to process the events in terms of the Markov condition and Markov-equivalence classes as they were to process the events in terms of low-level perceptual features and other cues.

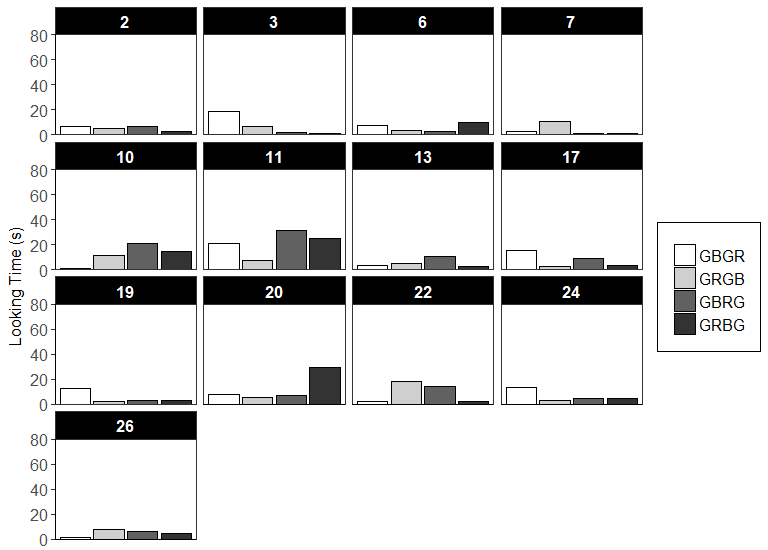


Fig. 22. Looking times to the four test events for the 13 “fast” habituators infants.

## Discussion

Experiment 5 was designed to examine whether infants between the ages of 18 and 22 months were sensitive to the Markov condition and Markov equivalence classes. The results revealed that infants in the present experiment did not show sensitivity to the Markov condition and Markov-equivalence classes. The results of the individual-difference analyses also revealed that younger and older infants as well as fast and slow habituators differed neither in their sensitivity to the Markov condition and Markov-equivalence classes nor did they differ in the extent to which they processed the test events in terms of the Markov condition, perceptual features, or other cues. Given that adults show sensitivity to the Markov condition and Markov-equivalence classes under some conditions (e.g., Experiment 1), the results of Experiment 5 indicate that it is not until after 22 months of age that infants develop sensitivity to the Markov condition and Markov equivalence classes and thus that the ability to represent causal events as CGMs emerges. The exact age at which infants develop such sensitivity is an issue that should be explored in future research.

# General Discussion

The main goals of the experiments presented here were to determine on what basis infants and adults process complex launching event sequences that consist of four objects and to determine whether both groups represent such sequences as CGMs based on their sensitivity to the Markov condition and Markov-equivalence classes or based on low-level perceptual features and cues. The results from Experiment 1 suggested that although adults are sensitive to the Markov condition and Markov-equivalence classes, such sensitivity was limited and appears to be driven by a small subset of subjects: adults showed sensitivity to the Markov condition and evidence that they formed Markov-equivalence classes when they were asked to provide causal comparisons but not when they were asked to provide perceptual comparisons. Experiment 2 extended these results to show that sensitivity to the Markov condition and Markov-equivalence classes did not hold in the presence of salient perceptual manipulations: adults perceived test events that included a novel spatial gap to be more inconsistent with the initial training event than test events that violated the Markov condition that was instantiated in the initial training event. Critically, compared to Experiment 1 in which a small proportion of subjects were shown to have processed the Markov condition and Markov-equivalence classes, only one adult processed the Markov condition and Markov-equivalence classes in Experiment 2.

Experiment 3 clarified the results of Experiment 2 by showing that the reason adults perceived test sequences that incorporated a spatial gap as more inconsistent than sequences without gaps was because these events were perceptually rather than causally distinct from the initial training event. The results from Experiment 4 demonstrated that adults were insensitive to, and did not selectively encode, the Markov condition when it was defined over either the shape or color dimension: adults perceived test sequences that included at least one perceptual change from initial training as more inconsistent than test sequences that violated the Markov condition in the training event—defined over the shape or color dimension—which suggested that the representation they formed of the training and generalized to the test sequences was grounded in low-level perceptual cues. Finally, the results from Experiment 5 indicated that infants between 18 and 22 months of age neither processed four-object three-chain launching sequences in terms of the Markov condition nor in terms of low-level perceptual features and cues. The results of a separate individual-difference analyses also indicated that there were no individual differences in how infants processed the test events; that is, neither younger or older infants nor fast or slow habituators processed the events in terms of the Markov condition and Markov-equivalence classes or in terms of low-level perceptual features.

Together, these results illustrate that sensitivity to the Markov condition and Markov-equivalence classes may not be as robust as has been thought and cast doubt on the claim that learners represent causal events as CGMs, which is a fundamental tenet of the Bayesian inference perspective (cf., Gopnik et al., 2004; Sobel & Kirkham, 2006). This conclusion is warranted by the fact that adults showed restricted sensitivity to both features and the fact that infants showed no sensitivity to both features. The similarities of these results notwithstanding, it is conceivable that these findings emerged for different reasons; that is, the mechanism or set of mechanisms that accounts for adults’ restricted sensitivity to the Markov condition and Markov-equivalence classes may differ from that that accounts for infants’ lack of sensitivity to the features. The aim of the next several sections will be to discuss this issue in some detail and to discuss the theoretical implications of the results for infants and adults. Given that different mechanisms are likely at play for infants and adults, the findings from the adult experiments and infant experiment—as well as their theoretical implications and limitations of the adult and infants experiments—will be discussed in separate sections.

## Summary of Adult Experiments

The results of the adult experiments have important implications for theories that claim that humans represent causal event as CGMs. One such theory that makes this claim—which is discussed in detail in Section 3.5 in the Introduction—is the Bayesian inference account (e.g., Gopnik et al., 2004). Recall that proponents of this perspective maintain that in the same way that humans possess an evolved visual system that reconstructs the visual environment from two-dimensional retinal input, so humans have an evolved causal system that reconstructs causal relations in the world as CGMs based on conditional independence information and causal data. As was discussed in the Introduction, the crux of this argument is that learners use a simple form of Bayes' rule to reason over a sample space that consists of potentially an infinite number of CGM structures and to choose which of these structures generated—and thus is most consistent with, and provides the best account of—the observed causal data. Support for this perspective derives from research that ostensibly shows that infants (e.g., Sobel & Kirkham, 2006, 2007), children (Gopnik et al., 2001; Sobel et al., 2004), and adults (e.g., Griffiths et al., 2011) are sensitive to and use Bayes’ rule to perceive and reason about causal events in a variety of causal contexts.

The finding that adults showed little evidence that they could represent four-object, three-chain launching sequences as CGMs is important because it challenges the fundamental tenet of the Bayesian inference account. This is because—unlike some of its less central claims such as the claim that humans use Bayes' rule to engage in backwards-blocking reasoning (e.g., Griffiths et al., 2011; Sobel et al., 2004; cf., Benton & Rakison, 2016)—representing causal events as CGMs is the *sine qua non* of this account. This is because the claim that humans use Bayes' rule to choose the most likely structure presupposes that causal events are represented as CGMs in the first place. In other words, a Bayesian analysis of causal events is only possible if the learner has represented such causal events as CGMs—in a potentially exponential space composed of CGMs—over which Bayes' rule can then operate to choose the structure with the highest posterior probability.

It is worth noting that not all proponents of CGMs believe that humans use a simple form of Bayes' rule to choose the CGM with the highest posterior probability. For example, Waldmann (1996, 2001) argued in his causal-models theory account that rather than using Bayes' rule to learn and choose the most likely causal structure, humans (a) compute conditional contrasts to determine the best causal model, (b) use temporal cues rather than conditional dependence information to learn the structure of CGMs, and (c) that these causal models are specialized and, as such, support domain-specific causal inferences. The results presented here have important implications for both the Bayesian-inference and causal-models accounts given that the ability to represent causal events as CGMs was limited: sensitivity to the Markov condition and Markov-equivalence classes was not robust to changes to low-level perceptual features and cues, emerged under some, but not under other, conditions, and was present in a relatively small subset of subjects. This means that both views should perhaps be endorsed cautiously as viable accounts of how humans represent causal events in the world.

Given that adults showed a greater tendency to respond to the events based on low-level perceptual features than in terms of either the Markov condition or Markov-equivalence classes, there is reason to think that they may have also processed the complex causal events used here associatively. This is because adults’ responses to the test events were consistent with what might be expected if they processed the events—and the relations between the objects and variables in these events—associatively. For example, adults in Experiment 1 could have only provided higher inconsistency ratings to the test events in which the blue and red objects were spatially contiguous if they recognized that each object was associated with a particular position in the sequence during training that was then violated in some of the test events but not in others. This account is consistent with the predictions of associative accounts (e.g., Rescorla & Wagner, 1972) and connectionist accounts (e.g., McClelland & Thompson, 2007), and it should be explored further in future research; that is, an important future direction will be to examine to what extent adults’ representations of the causal relation between objects and variables is grounded in the associative relation between the low-level perceptual features of such objects and variables. However, because the associative relations between the elements in the sequence and between their surface features were not experimentally or systematically manipulated, caution must be adopted before it can be concluded definitively that adults process associatively complex launching sequences.

Before concluding this section, it is worth mentioning three potential criticisms of the present studies. A first potential criticism is that the reason that adults’ sensitivity to the Markov condition and Markov-equivalence classes was restricted was because they were shown sequences that included temporal cues to causality; that is, adults in the present series of experiments explicitly were shown the order in which some objects caused other objects to move. These cues, in turn, could have directed adults' attention away from, rather than towards, the Markov condition in the sequences and caused them not to form Markov-equivalence classes. As Steyvers et al. (2003) argued, such cues may function to provide adults an "a priori bias about [causal] directionality" (p. 461) that precludes the need to represent complex causal events as CGMs. However, this criticism is weakened by the fact that the overall pattern of results in Experiment 1 suggested that adults were sensitive to the Markov condition and Markov equivalence classes and the fact that an analysis of individual differences in each of the four experiments revealed that a small, but nonetheless non-negligible, proportion of subjects showed sensitivity to the Markov condition and Markov-equivalence classes.

A second and related potential criticism concerns the nature of the perceptual manipulations that were made to the event sequences in Experiments 2 and 3; that is, a plausible explanation for why adults showed limited sensitivity to the Markov condition and Markov-equivalence classes was because the perceptual manipulations to the four-object launching sequences were so substantial that they disrupted this sensitivity. This, in turn, may have prevented adults from forming CGM representations of the event sequences. Although the present series of experiments do not offer a way definitively to rule out this possibility, note that proponents of the CGM accounts described above would maintain that sensitivity to the Markov condition and Markov-equivalence classes should hold in the presence of salient perceptual manipulations under some conditions (e.g., Gopnik et al., 2004; Gopnik & Tenenbaum, 2007). On this account, learners should show sensitivity to the Markov condition and Markov-equivalence classes in causal events that include salient perceptual manipulations so long as such manipulations do not alter the set of underlying conditional independencies (i.e., the Markov condition). Perhaps somewhat counterintuitively, some proponents of the CGM approach have even argued that certain salient perceptual manipulations may help, rather than hinder, learners’ sensitivity to the Markov condition, Markov-equivalence classes, and ability to form CGMs. For example, Waldmann (2001) and Fernbach and Sloman (2009) argued that information that pertains to temporal priority, temporal contiguity, and temporal chaining serve a facilitative, rather than an inhibitory, function in enabling adults to learn the structure of a causal graph and to distinguish between variables that are causes and variables that are recipients in these structures.

Such an account is also supported by the results of Experiment 1 that demonstrated that adults showed sensitivity to the Markov condition and Markov-equivalence classes despite the fact that temporal cues to causality were included in the training and test sequences. The finding from Experiment 1 would not have emerged if, as Steyvers et al. (2003) suggested, temporal cues to causality inhibited rather than facilitated adults’ sensitivity to the Markov condition and Markov-equivalence classes. The results from Experiments 1 to 4 provide still further support for the claim that temporal cues to causal structure facilitates sensitivity to the Markov condition and Markov-equivalence classes. In those experiments, limited sensitivity to the Markov condition and Markov-equivalence classes was observed even though some of the test events included a salient spatial gap (Experiments 2-3), temporal cues to causality (Experiments 1-4), and violations to other low-level perceptual features such as violations of either the color or shape of objects between training and test (Experiment 4).

A third potential criticism is that we did not examine whether, to what extent, and under what conditions adults show sensitivity to the Markov condition and whether they could form Markov-equivalence classes in non-launching event sequences. As mentioned in the Introduction, the reason we used launching-event sequences was because they are the simplest events in which to observe cause-and-effect relations. We acknowledge that by assessing sensitivity to the Markov condition and Markov-equivalence classes only in the context of launching-event sequences, we are necessarily limited in the kinds of inferences that can be made about sensitivity in non-launching contexts such as in common-effect and common-cause contexts. However, it is not possible to assess adults' sensitivity to the Markov condition and Markov-equivalence classes under common-effect structures because such structures form their own equivalence classes; that is, no two common-effect structures combine to form a single Markov-equivalence class. Given this fact, the number of available contexts in which to evaluate adults' sensitivity to Markov-equivalence classes is limited to common-cause and causal-chain structures. We chose to use causal-chain sequences rather than common-cause sequences in the present series of experiments because these sequences have been used extensively in previous research and the direction of causality between objects in these sequences arguably is simpler to discover than in common-cause sequences. It may nonetheless be worthwhile to assess adults’ sensitivity to the Markov condition and Markov-equivalence classes in the context of common-cause structures that include, rather than are bereft of, temporal cues to causality given that such cues appeared to have a facilitative effect under some conditions and for a small proportion of adults.

## Summary of Infant Experiments

As was discussed above, the results from Experiment 5 indicated that infants between the ages of 18 and 22 months of age showed no sensitivity to the Markov condition and Markov-equivalence classes. These results were interpreted to mean that sensitivity to these features does not emerge until sometime after 22 months of age. However, this interpretation must be accepted cautiously because there was also no indication that infants processed the events used here in terms of low-level perceptual features. Indeed, unlike adults who processed the causal events in terms of low-level perceptual features and cues, infants neither processed the events in terms of the Markov condition and Markov-equivalence classes nor did they process the events in terms of low-level features. This conclusion is based on the fact that their looking behavior was neither consistent with what would be predicted if they processed the events according to the Markov condition and Markov-equivalence classes nor with what would be predicted if they processed the events in terms of low-level perceptual features. Given these null findings, it is as yet unknown whether and to what extent a Bayesian inference mechanism or a domain-general learning mechanism underpins infants' processing of complex launching event sequences that are composed of four (or more) objects.

Open questions still remain about why sensitivity to the Markov condition and Markov-equivalence classes not present by 22 months of age? One possibility is that 18- to 22-month-old infants may not have processed causally the four-object, three-chain launching events, which may have indirectly prevented them from attending to the underlying conditional independencies; that is, perceiving the causality in launching events of various complexities may be a necessary, although not sufficient, prerequisite to encoding the Markov condition and Markov-equivalence classes and subsequently for representing causal events as CGMs. Because causality was not manipulated independently in Experiment 5, it is not yet possible to know whether the infants who showed sensitivity to the Markov condition and Markov-equivalence classes also perceived the events causally or whether the two features are perceived independently. Nonetheless, that there was evidence that some of the infants encoded the Markov condition and Markov-equivalence classes suggests that, at minimum, sensitivity to the Markov condition is not completely absent in infants between 18 and 22 months of age and may be a matter of individual differences based, perhaps, on being able to perceive causality in complex launching sequences.

Given that sensitivity to the Markov condition and Markov equivalence classes was shown not to be robust even by adulthood (Experiments 1-4), a second possibility is that such sensitivity may not emerge until well after the second year of life (cf., Sobel & Kirkham 2006, 2007). As discussed above, it may be the case that this sensitivity emerges only when infants can both perceive as well as reason about causality. Indeed, reasoning about causal events—through active intervention—may be an important prerequisite to encoding the Markov condition and Markov-equivalence classes. Indeed, successful causal reasoning requires that learners not only distinguish between causes and non-causes but recognize how causes are related through conditional probability in order to bring about different effects through causal action. Such causal action, then, coupled with the ability to perceive agents and recipients in causal sequences, may function to direct children's attention to the Markov condition and enable them to respond to sequences on the basis of their Markov equivalence. Given the dearth of research that has directly investigated this relation, this proposal is clearly speculative and represents a promising avenue for future research.

One promising avenue to gain traction on this proposal would be to examine whether infants show sensitivity to the Markov condition and Markov-equivalence classes under some conditions but not under others. It will be recalled that in Experiment 1 adults showed sensitivity to the Markov condition and were able to form Markov-equivalence classes when asked to provide causal similarity ratings but not when asked to provide perceptual similarity ratings. It may well be the case, then, that infants between the ages of 18 and 22 months will show sensitivity to these features if infants are made explicitly to attend to the causal relations, but not to the perceptual relations, in launching sequences such as those used here. For example, infants may encode the Markov condition and form equivalence classes if the salience of the impact between any two objects is increased, which can be achieved by playing an audible sound at the impact point.

If, in contrast, object salience does not direct learners' attention to conditional-independence relations (i.e., the Markov condition), then it is conceivable that infants would show no sensitivity to such relations if the salience of the surface features of particular objects in a causal sequence is increased. This can be achieved, for example, by either increasing the luminescence of particular objects in the sequence or by forcing particular objects in the sequence to expand and contract. These and similar manipulations—which are akin to asking adults to make causal and perceptual comparisons—would be useful in clarifying whether infants, like adults, show sensitivity to the Markov condition and Markov-equivalence classes under some but not other conditions. Note that this perspective assumes that sensitivity to the Markov condition and Markov-equivalence classes is present sometime between 18 and 22 months of age but will only be evinced under the right conditions and given the right manipulation or manipulations.

A third possibility is that because infants have accrued extensive causal experience by the second half of the second year of life, a process akin to "perceptual narrowing" (e.g., Kuhl et al., 2006; Pascalis, de Haan, & Nelson, 2002; Werker & Tees, 1984) for causal events may be at play that biases attention away from the Markov condition and Markov equivalence classes; that is, as infants exposure to causal events increases and their ability to perceive causality in launching sequences improves throughout early development, the ability to attend to the Markov condition and Markov-equivalence classes may suffer. This pattern of responding may emerge for two reasons. First, because an appreciation of cause-and-effect relations and the ability to generate novel effects through action is a cornerstone of the ability to understand the way the world works, overtime infants may learn selectively to attend to events on the basis of their causality rather than in terms of the conditional independence relations that underlie them. Second, this pattern of responding may emerge because the Markov condition and the causality of an event can sometimes have contravening implications. In such a case, learners would be forced to decide whether to respond to a causal event on the basis of causality or Markov-equivalence. In other words, two events can have different causal implications even if the two events are Markov equivalent. Consider the following two sequences:

1. A🡪B🡪C
2. C🡪B🡪A

Although the causal meanings of A, B, and C differ between the sequences—for example, A is the exogenous initiator in the first sequence, whereas C is the exogenous initiator in the second sequence—both sequences are Markov-equivalent; that is, both sequences instantiate identical conditional independencies. Infants who attend selectively to the Markov condition and respond to sequences on the basis of their Markov-equivalence would be expected to respond equivalently to both sequences. In contrast, infants who attend selectively to the causality of the sequences would be expected to treat both sequences differently, most especially after habituation to one of them. One plausible interpretation of infants' failure to attend to the Markov condition and Markov-equivalence classes in Experiment 5, then, is that these infants may be at a point in development in which sensitivity to the Markov condition and Markov-equivalence classes has diminished as their sensitivity to causality in the world heightens. However, because causality was not experimentally, systematically, or independently manipulated here, it is difficult to determine the viability of this account. Nonetheless, this will be an important issue to explore in future research. It will also be important to examine the viability of the "perceptual narrowing" assumption as it applies to causal events with infants younger than 18 months of age. This is because this account assumes that sensitivity to the Markov condition and Markov-equivalence classes may have sufficiently narrowed by the time infants are between 18 and 22 months of age.

It is worth noting that the speculative account outlined above has much in common with that forwarded by Rakison to explain infants’ developing knowledge of the motion properties of objects and entities and concept development more generally (Rakison & Lupyan, 2008). The crux of this perspective is that, because infants possess relatively impoverished information-processing abilities and have limited experience with motion events early in development, young infants will be restricted in the kinds of relations that they will encode. That is, initially infants will encode relations that involve simple static and dynamic features; yet, as their information-processing abilities continue to improve, infants will begin to encode relations that involve dynamic local and global features. Such a transition is thought to occur because dynamic features, unlike static features, require enhanced information-processing abilities to process due to their inherently more complex and dynamic nature and their intermittent availability in the perceptual array. Perhaps supporting this account, there is evidence that 10-month-olds will fail to encode the relation between static and dynamic features if those relations are embedded in a non-category discrimination context (e.g., Rakison & Poulin-Dubois, 2002); that 14-month-olds will fail to encode the same relations if embedded in a category context (e.g., Rakison, 2004); that 12-month-olds have no prior expectations about the kind of part (i.e., static or dynamic) that agents and recipients possess (e.g., Rakison, 2005); that 16-month-olds do not expect self-initiated objects to possess dynamic parts (e.g., Rakison, 2006); and that 14-month-olds attend to the structure and form of an object independently but not to the relation between them (Madole, Oakes, & Cohen, 1993).

However, as infants’ information-processing processing abilities continue to improve, they will be relatively unconstrained in the kinds of relations they will encode. Using a modified version of the launching-event paradigm, for example, Rakison (2005) found that 14-month-olds are as likely to associate a dynamic part with an agent and static part with a recipient as they are to associate a static part with an agent and a dynamic part with a recipient: following habituation either to a direct-launching event in which the agent possessed a dynamic part and the recipient possessed a static part or to a direct-launching event in which the agent possessed a static part and the recipient possessed a dynamic part, infants looked longer at the test events in which these part-object relations were reversed. Likewise, Rakison (2006) found that 18-month-olds, but neither 16-month-olds nor 20-month-olds, have not yet learned that objects that move without external contact tend to possess dynamic parts: following habituation to a noncausal (gap) event in which the first object possessed a static feature and the second object possessed a dynamic feature or the first object possessed a dynamic feature and the second object possessed a static feature, infants looked longer at the test event in which these part-object relations were violated. This finding is particularly surprising because if infants at 18 months of age recognized that objects that move without external contact tend to possess dynamic parts, they should have only shown elevated looking when the second object possessed a static part but when it possessed a dynamic part.

With still more experience with motion events and as their experience with the regularities in the world accrues, infants will tend to encode those relations that make sense in the world, for which relational correlations are the strongest (e.g., the correlation between legs and walking), and that are consistent with their prior experience. Support for this contention can also be found in the studies cited above that show that 22-month-olds can readily encode the relation between dynamic local features and global features to represent objects (e.g., Rakison & Poulin-Dubois, 2002) and categories of objects (e.g., Rakison, 2004), that 16-month-olds expect agents to possess dynamic features and recipients to possess static features (e.g., Rakison, 2005), that 20-month-olds expect self-propelled objects to possess dynamic features (e.g., Rakison, 2006), and that 20-month-olds recognize that the function of an object is related to its structure (e.g., Madole et al., 1993).

As is implied by this account, this developmental progression is thought to be the result of an ever-improving information-processing system (e.g., Cohen & Cashon, 2003), a sensitive perceptual system that is biased to attend to dynamic information (e.g., Dannemiller, 2000), and a associative mechanism that becomes increasingly constrained in the kinds of relations that it encodes to reflect those found in the real world (for an extensive review of this account see Rakison & Lupyan, 2008).

This perspective can be extended to explain the results of Experiment 5 and the development of causal perception more generally. Early on infants will be unable to perceive the causality in launching event sequences, much less be able to encode the Markov condition and form Markov-equivalence classes, due to impoverished information-processing abilities and a lack of real-world causal experience. Yet, because infants are thought to possess a sensitive perceptual system—or inherent perceptual biases—that orient their attention to motion, they will display a preference for the basic motion properties of launching sequences. This may explain why infants at birth become entrained by (i.e., prefer) the continuous, smooth motion of a direct-launching (e.g., Mascalzoni et al., 2013) and continue to do so at 4 months of age (e.g., Cohen & Amsel, 1998) if they are not given direct causal action experience (e.g., Rakison & Krogh, 2012). However, as infants' information-processing abilities continue to improve but before they have sufficiently encoded the regularities in the world, they should be relatively unconstrained in the relations they represent in causal events. This may explain why, at 5½ months of age, infants distinguish equally between the direct (causal) launching event and non-causal events and have little difficulty encoding the spatiotemporal relations between the objects in such events (e.g., Cohen & Amsel, 1998; Leslie, 1984). Only when infants' information-processing abilities improve still more and they become sensitive to the regularities that govern causal events in the real world would they be expected to organize causal events on the basis of their causality. This accounts for why infants can perceive causality in launching events and can distinguish noncausal events as a single category that is distinct from the direct (causal) launching event at 6½ months of age for artificial stimuli (e.g., Cohen & Amsel, 1998; Leslie & Keeble, 1987) and at 10 months of age for 3D realistic stimuli (Oakes & Cohen, 1990).

The developmental progression of causal perception likely does not stop there, given that infants in the present experiment failed to show sensitivity to the Markov condition and Markov equivalence classes when shown complex launching event sequences. The reason infants may have failed, then, to encode the Markov condition and form Markov-equivalence classes is because by 18 months of age they learn that successfully bringing about new effects through causal intervention is more related to understanding how a particular causal system works and less related to the pattern of underlying conditional independencies that typify the Markov condition. This is especially true given that, as was mentioned above, two (or more) events can have different causal implications despite the fact that they may encode the same pattern of conditional independencies. Note that this account and the “perceptual narrowing” account described above similarly predict that sensitivity to the Markov condition lessens with age. However, both accounts differ in terms of the proposed underlying mechanism for why infants between the ages of 18 and 22 months fail to encode the Markov condition and form Markov-equivalence classes.

Because Experiment 5 was not designed to examine to what extent infants between 18 and 22 months of age process four-object launching events associatively, it is difficult to know the viability of this account in the context of Experiment 5. In addition, because infants younger than 18 months of age were not tested, it is an open question whether infants' sensitivity, like their sensitivity to causality, undergoes a developmental progression. It will be important to test infants who are younger than those tested in Experiment 5 but who can process causal chain sequences. In addition, it will be important systematically to manipulate the associative relations between objects and features to determine to what extent a domain-general associative learning mechanism underpins infants' learning of complex causal chain sequences and to determine in what ways this learning is constrained by prior experience and real-world causal regularities.

Based on the foregoing discussion, more research is clearly needed to determine which of these alternative explanations best accounts for infants' processing of the Markov condition and Markov-equivalence classes in complex causal event sequences that consist of multiple objects. Until this research is carried out, caution must be adopted before strong claims can be made about whether, to what extent, and in what ways a Bayesian-inference or a domain-general mechanism underpins infants' processing of complex event sequences, although there may be reason to believe that domain-general associative learning, but not Bayesian inference, underpins infants' processing of causal events. This is based on the fact that adults in the present series of experiments processed the causal events in terms of low-level perceptual features, the fact that a great deal of research on infant causal perception suggests that domain-general associative learning, rather than innate modules (e.g., Leslie, 1994, 1995), mechanisms (e.g., Mandler, 1992), skeletal systems (Gelman, 1990), or core systems (Spelke & Kinzler, 2007), may underpin infants' perception of causal events (e.g., Oakes & Cohen, 1990; Cohen, Chaput, & Cashon, 2002; Rakison, 2005, 2006; Rakison, Smith, & Ali, 2016).

Notwithstanding these alternative explanations for the present data, the results of Experiment 5 indicate that infants between 18 and 22 months of age are unable to encode the Markov condition and form Markov-equivalence classes with the current set of stimuli and under the current testing conditions. Although one must exercise caution before it can be claimed that these results constitute conclusive evidence against the Bayesian-inference perspective—because, as was just discussed, the null findings could have arisen for a number of reasons—at the very least these results challenge the strong claim of the Bayesian-inference perspective that humans possess a causal system that is operative from birth (or shortly thereafter) that enables them to encode the Markov condition, form Markov-equivalence classes, and, ultimately, to represent causal events as CGMs (e.g., Gopnik et al., 2004; Gopnik & Wellman, 2012).

## Conclusion

In summary, the results presented here demonstrate that although adults are sensitive to the Markov condition and Markov-equivalence classes, this sensitivity may not be robust as was originally thought based on the fact that it was evinced only under specific conditions, is disrupted by salient perceptual manipulations, and may be grounded in low-level perceptual cues and features. To my knowledge, this is the first series of experiments to use four-object, rather than two- or three-object, sequences to examine on what basis infants between 18 and 22 months of age and adults process such sequences and whether, to what extent, and under what conditions both groups show sensitivity to the Markov condition and Markov-equivalence classes and evidence that they represent causal events as CGMs. Given that it was the case that sensitivity to the Markov condition and Markov-equivalence classes was limited (Experiments 1-4) or nonexistent (Experiment 5), a potentially fruitful avenue of future research will be to examine under what other conditions will infants and adults show sensitivity to the Markov condition and Markov-equivalence classes and whether such sensitivity can be facilitated in other causal contexts that include, but importantly are not limited to, four-object launching-event sequences.

References

Allan, L. G., & Jenkins, H. M. (1980). The judgment of contingency and the nature of the response alternatives. Canadian Journal of Psychology/Revue canadienne de psychologie, 34(1), 1.

Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. Learning and Motivation, 14(4), 381-405.

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser?. Journal of experimental psychology: General, 108(4), 441-485.

Baillargeon, R. (1987). Object permanence in 3½-and 4½-month-old infants. Developmental psychology, 23(5), 655.

Bélanger, N. D., & Desrochers, S. (2001). Can 6‐month‐old infants process causality in different types of causal events?. British Journal of Developmental Psychology, 19(1), 11-21.

Benton. D.T., & Rakison, D.H. Causal Reasoning in Infants and Adults: Revisiting backwards-blocking. Proceedings of the Annual Conference of the Cognitive Society, 38, pp. 830-835.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of memory and language, 59(4), 390-412.

Baron-Cohen, S. E., Tager-Flusberg, H. E., & Cohen, D. J. (1994). Understanding other minds: Perspectives from autism. In Most of the chapters in this book were presented in draft form at a workshop in Seattle, Apr 1991. Oxford University Press.

Bélanger, N. D., & Desrochers, S. (2001). Can 6‐month‐old infants process causality in different types of causal events?. British Journal of Developmental Psychology, 19(1), 11-21.

Blaisdell, A. P., Bristol, A. S., Gunther, L. M., & Miller, R. R. (1998). Overshadowing and latent inhibition counteract each other: Support for the comparator hypothesis. Journal of Experimental Psychology: Animal Behavior Processes, 24(3), 335.

Blaser, R. E., Couvillon, P. A., & Bitterman, M. E. (2004). Backward blocking in honeybees. Quarterly Journal of Experimental Psychology Section B, 57(4), 349-360.

Bonawitz, E., Denison, S., Griffiths, T. L., & Gopnik, A. (2014). Probabilistic models, learning algorithms, and response variability: Sampling in cognitive development. Trends in cognitive sciences, 18(10), 497-500.

Buehner, M. J., & Humphreys, G. R. (2010). Causal contraction: spatial binding in the perception of collision events. Psychological Science, 21(1), 44-48.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. The developmental psychology of time, 209-254.

Carroll, C. D., Cheng, P. W., & Lu, H. (2013). Inferential dependencies in causal inference: A comparison of belief-distribution and associative approaches. Journal of Experimental Psychology: General, 142(3), 845.

Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgment. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(5), 837.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. Journal of cognitive neuroscience, 10(1), 1-34.

Choi, H., & Scholl, B. J. (2006). Perceiving causality after the fact: Postdiction in the temporal dynamics of causal perception. Perception, 35(3), 385-399.

Chowdhry, S. (2011). CRITIQUE OF MANDLER’S THEORY OF PERCEPTUAL ANALYSIS. Journal of Cognitive Science, 6.

Cheng, P. W. (1997). From covariation to causation: A causal power theory.Psychological review, 104(2), 367-405.

Clifton Jr, C., Staub, A., & Clifton, C. (2011). Syntactic influences on eye movements during reading. eye, 3(2).

Cohen, L. B. (1973). A two process model of infant visual attention. Merrill-Palmer Quarterly of Behavior and Development, 19(3), 157-180.

Cohen, L. B. (1988). An information-processing approach to infant cognitive development.

Cohen, L. B., Amsel, G., Redford, M. A., & Casasola, M. (1998). The development of infant causal perception. Perceptual development: Visual, auditory, and speech perception in infancy, 167-209.

Cohen, L. B., & Cashon, C. H. (2003). Infant perception and cognition. Handbook of psychology: Developmental psychology, 6, 65-89.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). Habit 2000: A new program for obtaining and organizing data in infant perception and cognition studies. Version, 1, 2000-2002.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. Cognitive Development, 17(3-4), 1323-1343.

Cohen, L. B., Rundell, L. J., Spellman, B. A., & Cashon, C. H. (1999). Infants' perception of causal chains. Psychological Science, 10(5), 412-418.

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. Journal of experimental psychology: Learning, memory, and cognition, 35(3), 678.

Danks, D. (2003). Equilibria of the Rescorla–Wagner model. Journal of Mathematical Psychology, 47(2), 109-121.

Danks, D. (2009). The psychology of causal perception and reasoning. Oxford: Oxford University Press.

Dannemiller, J. L. (2000). Competition in early exogenous orienting between 7 and 21 weeks. Journal of Experimental Child Psychology, 76(4), 253-274.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. Quarterly Journal of Experimental Psychology, 49B, 60–80.

Denison, S., Bonawitz, E. B., Gopnik, A., & Griffiths, T. L. (2010, January). Preschoolers sample from probability distributions. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 32, No. 32).

Dwyer, D. M., Mackintosh, N. J., & Boakes, R. A. (1998). Simultaneous activation of the representations of absent cues results in the formation of an excitatory association between them. Journal of Experimental Psychology: Animal Behavior Processes, 24(2), 163.

Erlick, D. E., & Mills, R. G. (1967). Perceptual quantification of conditional dependency. Journal of Experimental Psychology, 73(1), 9.

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. Journal of experimental psychology: Learning, memory, and cognition, 35(3), 678.

Fodor, J. A. (1985). Precis of the modularity of mind. Behavioral and brain sciences, 8(1), 1-5.

Fonlupt, P. (2003). Perception and judgement of physical causality involve different brain structures. Cognitive Brain Research, 17(2), 248-254.

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. Cognition, 117(2), 107-125.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate‐inanimate distinction as examples. Cognitive science, 14(1), 79-106.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. Trends in cognitive sciences, 7(7), 287-292.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. Journal of Experimental Psychology: General, 117(3), 227.

Gluck, M. A., & Bower, G. H. (1990). Component and pattern information in adaptive networks.

Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. Developmental psychology, 37(5), 620.

Gopnik, A. (2003). The theory theory as an alternative to the innateness hypothesis. Chomsky and his critics, 238-254.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. Psychological review, 111(1), 3.

Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. Trends in cognitive sciences, 8(8), 371-377.

Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. Trends in cognitive sciences, 8(8), 371-377.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. Child development, 71(5), 1205-1222.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. Developmental psychology, 37(5), 620.

Gopnik, A., & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. Developmental science, 10(3), 281-287.

Gopnik, A., & Wellman, H. M. (1994). 10 The theory theory. Mapping the mind: Domain specificity in cognition and culture, 257.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. Psychological bulletin, 138(6), 1085.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. Cognitive Science, 35(8), 1407-1455.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. Memory & Cognition, 30(7), 1128-1137.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. Cognition, 61(3), 233-259.

Houwer, J. D., & Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. The Quarterly Journal of Experimental Psychology: Section B, 55(4), 289-310.

Hume, D. (1748). Philosophical Essays Concerning Human Understanding: By the Author of the Essays Moral and Political. A. Millar.

Hume, D. (1777/1993). An Enquiry Concerning Human Understanding. Indianapolis:Hackett.

Humphreys, G. W., & Forde, E. M. (2001). Hierarchies, similarity, and interactivity in object recognition:“Category-specific” neuropsychological deficits. Behavioral and Brain Sciences, 24(3), 453-476.

Jeffreys, H. (1961). Theory of probability (3rd Ed.). Oxford, UK: Oxford University Press.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. Psychological Monographs: General and Applied, 79(1), 1.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. Punishment and aversive behavior.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: rethinking the study of developmental disorders. Developmental psychology, 45(1), 56.

Karmiloff-Smith, A., Scerif, G., & Ansari, D. (2003). Double dissociations in developmental disorders? Theoretically misconceived, empirically dubious. Cortex, 39(1), 161-163.

Keil, F. C. (1981). Constraints on knowledge and cognitive development. Psychological review, 88(3), 197.

Kestenbaum, R., Termine, N., & Spelke, E. S. (1987). Perception of objects and object boundaries by 3‐month‐old infants. British journal of developmental psychology, 5(4), 367-383.

Kloos, H., & Sloutsky, V. M. (2013). Blocking a redundant cue: what does it say about preschoolers' causal competence?. Developmental science, 16(5), 713-727.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Developmental science, 9(2), F13-F21.

Kuhlmeier, V. A., Bloom, P., & Wynn, K. (2004). Do 5-month-old infants see humans as material objects?. Cognition, 94(1), 95-103.

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. Developmental psychology, 43(1), 186.

Lagnado, D. A., & Sloman, S. (2004). The advantage of timely intervention. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(4), 856.

Larkin, M. J., Aitken, M. R., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies.Journal of Experimental Psychology: Learning, Memory, and Cognition,24(6), 1331-1352.

Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. Perception, 13(3), 287-305.

Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. Mapping the mind: Domain specificity in cognition and culture, 119-148.

Leslie, A. M. (1995). A theory of agency. Causal cognition: A multidisciplinary debate, 121-141.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality?. Cognition, 25(3), 265-288.

Lewis, D. (1973). Counterfactuals and comparative possibility. In Ifs (pp. 57-85). Springer, Dordrecht.

Lewis, D. (1986). On the plurality of worlds. Oxford, 14, 43.

Lovibond, P. F., Been, S., Mitchell, C. J., Bouton, M. E., & Frohardt, R. (2003). Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. Memory & Cognition, 31, 133–142.

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. Cognition, 131(2), 284-299.

Lupyan, G., & Rakison, D. H. (2006, January). What moves in a mysterious way? A domain-general account of learning about animacy and causality. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 28, No. 28).

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. Psychological review, 99(4), 587.

Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infancy. Cognition, 59(3), 307-335.

Madole, K. L., & Oakes, L. M. (1999). Making sense of infant categorization: Stable processes and changing representations. Developmental Review, 19(2), 263-296.

Madole, K. L., Oakes, L. M., & Cohen, L. B. (1993). Developmental changes in infants' attention to function and form-function correlations. Cognitive Development, 8(2), 189-209.

Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: newborns’ preference for physical causality. Developmental science, 16(3), 327-335.

McClelland, J. L., & Thompson, R. M. (2007). Using domain‐general principles to explain children's causal reasoning abilities. Developmental Science, 10(3), 333-356.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. Science, 198(4312), 75-78.

Melz, E. R., Cheng, P. W., Holyoak, K. J., & Waldmann, M. R. (1993). Cue competition in human categorization: Contingency or the Rescorla-Wagner Learning Rule? Comment on Shanks (1991).

Michotte, A. (1946). La Perception de la Causalite (Louvain: Institut Supe rieur de Philosophie).

Michotte, A. E. (1963). The Perception of Causality. (Tr. byT.R.Miles and E. Miles.). London: Methuen. Original work published 1946.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. Psychological bulletin, 117(3), 363.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. Journal of Personality and Social psychology, 67(6), 949.

Moors, P., Wagemans, J., & de-Wit, L. (2017). Causal events enter awareness faster than non-causal events. PeerJ, 5, e2932.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences, 23(3), 299-325.

Needham, A. (1999). How infants grasp two adjacent objects: Effects of perceived display composition on infants’ actions. Developmental Science, 2(2), 219-233.

Newman, G. E., Choi, H., Wynn, K., & Scholl, B. J. (2008). The origins of causal perception: Evidence from postdictive processing in infancy. Cognitive psychology, 57(3), 262-291.

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. Cognitive Development, 5(2), 193-207.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life?. Science, 296(5571), 1321-1323.

Piaget, J. (1952). The construction of reality in the child. New York: Basic Books.

Povinelli, D. (2000). Folk physics for apes: The chimpanzee's theory of how the world works.

R Development Core Team (2008). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URLhttp://www.R-project.org.

Rakison, D. H. (2004). Infants' sensitivity to correlations between static and dynamic features in a category context. Journal of Experimental Child Psychology, 89(1), 1-30.

Rakison, D. H. (2005). A secret agent? How infants learn about the identity of objects in a causal scene. Journal of Experimental Child Psychology, 91(4), 271-296.

Rakison, D. H. (2006). Make the first move: How infants learn about self-propelled objects. Developmental Psychology, 42(5), 900.

Rakison, D. H., & Krogh, L. (2012). Does causal action facilitate causal perception in infants younger than 6 months of age?. Developmental science, 15(1), 43-53.

Rakison, D. H., & Lupyan, G. (2008). Developing object concepts in infancy: An associative learning perspective. Monographs of the Society for Research in Child Development, i-127.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate–inanimate distinction. Psychological bulletin, 127(2), 209.

Rakison, D. H., & Poulin‐Dubois, D. (2002). You go this way and I’ll go that way: Developmental changes in infants’ detection of correlations among static and dynamic features in motion events. Child Development, 73(3), 682-699.

Rakison, D. H., Smith, G. T., & Ali, A. (2016). Who is the dynamic duo? How infants learn about the identity of objects in a causal chain. Developmental psychology, 52(3), 355.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory, 2, 64-99.

Rakison, D. H., & Yermolayeva, Y. (2011). How to identify a domain-general learning mechanism when you see one. Journal of Cognition and Development, 12(2), 134-153.

Raftery, A. E. (1995). Bayesian model selection in social research. Sociological methodology, 111-163.

Rice, M. L., Warren, S. F., & Betz, S. K. (2005). Language symptoms of developmental language disorders: An overview of autism, Down syndrome, fragile X, specific language impairment, and Williams syndrome. Applied psycholinguistics, 26(1), 7-27.

Rogers, T. T., & McClelland, J. L. (2004). Semantic cognition: A parallel distributed processing approach. MIT press.

Rovee-Collier, C., & Cuevas, K. (2008). The development of infant memory. In The development of memory in infancy and childhood (pp. 23-54). Psychology Press.

Schlottmann, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. The Quarterly Journal of Experimental Psychology, 44(2), 321-342.

Schlottmann, A., Cole, K., Watts, R., & White, M. (2013). Domain-specific perceptual causality in children depends on the spatio-temporal configuration, not motion onset. Frontiers in psychology, 4, 365.

Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events. Psychological Science, 13(6), 493-498.

Scholl, B. J., & Nakayama, K. (2004). Illusory causal crescents: Misperceived spatial relations due to perceived causality. Perception, 33(4), 455-469.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. Trends in cognitive sciences, 4(8), 299-309.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. The Quarterly Journal of Experimental Psychology Section B, 37(1b), 1-21.

Shanks, D. R. (1995). The psychology of associative learning (Vol. 13). Cambridge University Press.

Shultz, T. R. (2003). Computational developmental psychology. Cambridge, MA: The MIT Press

Shanks, D. R., & Dickinson, A. (1988). Associative accounts of causality judgment. In Psychology of learning and motivation (Vol. 21, pp. 229-261). Academic Press.

Smedslund, J. (1963). The concept of correlation in adults. Scandinavian Journal of Psychology, 4(1), 165-173.

Sobel, D. M. (2004). Exploring the coherence of young children's explanatory abilities: Evidence from generating counterfactuals. British Journal of Developmental Psychology, 22(1), 37-58.

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: the development of causal reasoning in toddlers and infants. Developmental psychology, 42(6), 1103.

Sobel, D. M., & Kirkham, N. Z. (2007). Bayes nets and babies: Infants’ developing statistical reasoning abilities and their representation of causal knowledge. Developmental science, 10(3), 298-306.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children’s causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. Cognitive science, 28(3), 303-333.

Spelke, E. S. (1990). Principles of object perception. Cognitive science, 14(1), 29-56.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. Developmental science, 10(1), 89-96.

Spelke, E. S., & Kinzler, K. D. (2009). Innateness, learning, and rationality. Child development perspectives, 3(2), 96-98.

Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action.

Spelke, E. S., & Newport, E. L. (1998). Nativism, empiricism, and the development of knowledge.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. Cognitive science, 27(3), 453-489.

Sugarman, S. (1987). The priority of description in developmental psychology. International Journal of Behavioral Development, 10(4), 391-414.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. Psychological review, 88(2), 135.

Taylor, A. H., Cheke, L. G., Waismeyer, A., Meltzoff, A. N., Miller, R., Gopnik, A., ... & Gray, R. D. (2014). Of babies and birds: complex tool behaviours are not sufficient for the evolution of the ability to create a novel causal intervention. Proc. R. Soc. B, 281(1787), 20140837.

Tomasello, M., & Call, J. (1997). Primate cognition. Oxford University Press, USA.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. Learning and motivation.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems ofp values. Psychonomic bulletin & review, 14(5), 779-804.

Wagner, A. R., Mazur, J. E., Donegan, N. H., & Pfautz, P. L. (1980). Evaluation of blocking and conditioned inhibition to a CS signaling a decrease in US intensity. Journal of Experimental Psychology: Animal Behavior Processes, 6(4), 376.

Waldmann, M. R. (1996). Knowledge-based causal induction. Psychology of Learning and Motivation, 34, 47-88.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(1), 53.

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. Psychonomic Bulletin & Review, 8(3), 600-608.

Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. Cognition, 133(2), 343-357.

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(1), 174.

Wasserman, E. A., & Shaklee, H. (1984). Judging response-outcome relations: The role of response-outcome contingency, outcome probability, and method of information presentation. Memory & Cognition, 12(3), 270-286.

Warrington, E. K. (1981). Neuropsychological studies of verbal semantic systems. Phil. Trans. R. Soc. Lond. B, 295(1077), 411-423.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. Developmental psychology, 34(6), 1289.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant behavior and development, 7(1), 49-63.

White, P. A. (1988). Causal processing: Origins and development. Psychological Bulletin, 104(1), 36.

White, P., & Milne, A. (2003). Visual impressions of penetration in the perception of objects in motion. Visual Cognition, 10(5), 605-619.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. Cognition, 69(1), 1-34.

Younger, B. A., & Cohen, L. B. (1983). Infant perception of correlations among attributes. Child Development, 858-867.

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. Child development, 803-815.

1. It is important to note that although, in general, Hume believed that knowledge derives from direct experience, it is still possible to predict one event from another based on experiencing similar events in the past, a principle that Hume terms "equivalence." This is an important distinction to make given that most theorists assume, and incorrectly argue, that Hume believed strictly that humans have to experience a specific correlated event in order to learn about the causality of it. [↑](#footnote-ref-1)
2. It should be noted that this model is similar to the Widrow-Hoff (delta) rule that is used in connectionist computational models of learning (e.g., Shanks, 1985); in many cases, the RW rule converges to the Widrow-Hoff rule in the limit and for certain parameters (e.g., Danks, 2003; Gluck & Bower, 1988, 1990; Melz, Cheng, Holyoak, & Waldmann, 1993; Sutton & Barto, 1981). [↑](#footnote-ref-2)
3. Note that two graphs are said also to form a Markov-equivalence class if they can be expressed in terms of identical joint-probability distributions. For example, the two graphs in Figure 1C can be expressed as *P*(*B*)*P*(*A*|*B*)*P*(*C*|*B*). [↑](#footnote-ref-3)