Spring 2019

"She Sneezed and the Germs Went Onto Him!": Children's Conception of Probabilistic Causality in Illness

Seetha Davis , '19

Recommended Citation

Davis, Seetha , '19, ""She Sneezed and the Germs Went Onto Him!": Children's Conception of Probabilistic Causality in Illness" (2019). *Senior Theses, Projects, and Awards*. 363.
https://works.swarthmore.edu/theses/363

“She Sneezed and the Germs Went Onto Him!”: Children’s Conception of Probabilistic Causality in Illness

Seetha Davis

Swarthmore College

May 2019

Abstract

Understanding sickness and engaging in preventative health behaviors necessitate abilities to reason causally and probabilistically. This set of studies addresses these cognitive processes in young children. Prior literature suggests that four-year-olds see causes of disease as determined (e.g., germ exposure necessarily causes illness) and cannot distinguish between causes of disease in terms of differing degrees of contagion (e.g., a germ-induced cough is equally contagious as a smoke-induced cough). In order to assess young children’s probabilistic and causal reasoning abilities in the health domain, two studies were conducted. Study 1 investigated probabilistic thinking in adults and children in the health domain. Participants were presented with stories that either did or did not include preventative health behavior (e.g., washing one’s hands to prevent spread of illness). The results suggested that, when presented with stories both containing and not containing preventative health behaviors, children can think probabilistically about illness transmission. Study 2 evaluated the utility of cognitive comparison in supporting children’s causal reasoning. In contrast to prior literature, a novel comparison learning task facilitated children’s understanding of different levels of contagion associated with various causes of illness. The results suggested that, overall, comparing two stories involving different causes of illness and different contagion outcomes facilitates differentiation between the level of contagion for various causes of illness. Taken together, these findings substantiate the utility of comparison learning in supporting young children to think about probabilistic causality in the health domain.

We are bombarded with messages that remind us to wash our hands. In public restrooms, we see, “Employees must wash hands before returning to work.” In classrooms, colorful posters proclaim, “Wash your hands!” with illustrations detailing the appropriate songs to sing for the correct duration of lather time. However, very few of these posters or placards explain *why* handwashing is important; we assume that everyone knows that washing one’s hands fights germs and bacteria, and germs and bacteria cause illness, so washing hands will prevent illness. If one were to ask a preschooler why handwashing is important, the child would likely say that people wash their hands so that they do not get sick. But there is little evidence to indicate that children can connect handwashing to a larger understanding of what causes sickness. In other words, it is unclear whether children understand that handwashing targets the illness-causing germs; we do not know if they comprehend this multi-step causal mechanism. Complete and correct causal understanding is likely linked to engagement in (preventative) health behaviors. Performance of health behaviors in childhood predicts both health behaviors and health outcomes in adulthood (Kjønniksen, Torsheim, & Wold, 2008; Paavola, Vartiainen, & Haukkala, 2004). Thus, in order to understand why people do or do not engage in health habits, it is critical to understand the ways in which children think about cause-and-effect relationships in the health domain. This work sought to investigate young children’s health-related causal understanding both through further characterization of their difficulty in the health domain and through exploring potential ways to foster this understanding.

Causal reasoning describes the ability to identify and understand the relationship of cause-and-effect; this type of reasoning can be applied to many domains. Past studies, for example, have examined it with regard to concepts of gravity and collisions; in these physical domains, even infants understand fundamental causal concepts (Göksun, George, Hirsh-Pasek, &

Golinkoff, 2013; Leslie & Keeble, 1987). In the health domain, causal understanding is necessary in order to engage in appropriate health behaviors – both preventative and treatment-oriented behaviors; one will not engage in any kind of health behavior unless the behavior is linked to (lack of) illness. Understanding the ways in which diseases are contracted and how health behavior disrupts disease contraction are fundamentally examples of understanding cause-and-effect relationships.

For example, bacteria cause sickness, which might manifest as a cough. That series of cause and effects – bacteria-(sickness)-cough – can be interrupted by washing one’s hands, which establishes a new cause-and-effect relationship: bacteria-handwashing-(no sickness)-no cough. In order to understand the consequences of a disease and the impact of a preventative health behavior such as handwashing on that disease, one must be able to reason about the cause-and-effect relationships at play. Although children can recognize and apply complex causal concepts in other domains (e.g., Göksun et al., 2013), preschool-age children do not reason accurately about disease causation and contagion (Solomon & Cassimatis, 1999). The root of this inability may lie in the challenge inherent in understanding causality in the health domain.

Causal reasoning about health is uniquely difficult because health-related causes are not determined. Exposure to germs does not necessarily make one sick, although germ-exposure without handwashing certainly increases one’s likelihood of contracting an illness. Thus, understanding causality alone only gets one so far in predicting health outcomes. Full understanding requires *probabilistic causality*. There are two realms of causality: deterministic and probabilistic. Deterministic causal relations are certain; if one drops a ball from shoulder height, it will fall to the ground. Probabilistic causal relations, on the other hand, are uncertain; it might be likely that eating half of a pizza will give one a stomachache, but such a relationship is not guaranteed. Such probabilistic reasoning still requires a specific understanding of cause and effect; one will not expect eating half of a pizza to induce hearing loss.

Probabilistic reasoning is the ability to reason about outcomes from a hypothesis, with understandings of likelihoods of those outcomes (Wedell, 2011). In the health domain, probabilistic reasoning is most salient in tasks involving risk assessment: How likely is it that I will get sick if I choose to wash my hands? How does that likelihood change if I choose not to wash my hands? Questions such as these necessitate several skills, including abilities to assess statistical input and generate probabilistic hypotheses; use evidence to inform decision-making; and generate risk assessments associated with engaging in certain behaviors over others.

Although the combination of these abilities may seem complex, many of the causal relationships we know and understand (e.g., emotional causal relations such as the likely impact of scolding or complimenting someone) are probabilistic. However, most studies, especially those with children, explore only deterministic causality; few studies have investigated probabilistic causal relationships with young children, despite their prevalence. One study suggests that four-year-olds are capable of making inferences about causal strength based on probabilistic information. Children were presented with a light on a box that lit up only when the child or the experimenter performed certain actions and were asked to figure out which actions triggered illumination. Children reported stronger causal relationships when illumination frequency was higher, suggesting that four-year-olds use information about frequency (i.e., statistical input) to infer strength of causal relationships (Kushnir & Gopnik, 2005).

By contrast, another study exploring probabilistic causality in the health domain – through stories about children exposed to illness irritants – concluded that preschool-age children have some probabilistic capacities but do not extend them to the health domain. There, young children were told stories about illness transmission and then asked to guess whether the protagonist would get sick and provide a judgment of their certainty for their sickness prediction (i.e., “for sure” or “just maybe”). For stories related to illness, five-year-olds were overwhelmingly certain in their predictions, despite being appropriately uncertain for non-illness-related stories detailing probabilistic causal relationships such as a child choosing a snack without prior demonstrated preferences (Kalish, 1998).

Full understanding of health and illness is difficult because it requires both probabilistic and causal reasoning, along with the synthesis of those reasoning skills into the development of probabilistic causality. Because health behaviors – and habits, in general – are most often developed in childhood, this series of studies seeks to explore probabilistic and causal reasoning in children in order to understand why reasoning in the health domain is challenging for preschoolers, i.e., three- to six-year-olds.

This thesis presents two studies. The first study, involving both children and adults, investigated probabilistic reasoning in the health domain with respect to engagement in preventative health behaviors such as handwashing. The second study investigated whether a specific learning strategy – comparison – can aid children’s probabilistic causal understanding in the health domain. All of the following studies were approved by the Swarthmore College Institutional Review Board (Protocol 12-13-041).

Part I: Probabilistic Understanding

Engaging in any form of preventative health behavior requires probabilistic understanding. Prior research suggests that adults have fairly advanced probabilistic reasoning skills (Evans, Handley, Perham, Over, & Thompson, 2000), despite sometimes using faulty heuristics and faulty shortcuts (Tversky & Kahneman, 1974).

Children can also think probabilistically; research suggests the roots of probabilistic reasoning develop in infancy. Findings from a series of studies indicate that children as young as six months old can make probabilistic inferences. Given two boxes, one with mostly pink and the other with mostly yellow balls, infants looked longer when a human hand produced the improbable ball, i.e., a yellow ball from the box with mostly pink balls (Denison, Reed, & Xu, 2013). Another study with 12-month-old children concluded that infants possess complex probabilistic reasoning skills when presented with complex object movements, consistent with a Bayesian observer model. For example, infants were shown multiple objects moving in a rectangular space with three exits on one side and one exit on the other. The display was then occluded, and one object came out from behind the screen. Twelve-month-old children looked longer when the emerging object came from the one-exit side (versus the three-exit side). These findings suggest that 12-month-olds make probabilistic inferences that inform their expectations about physical movement (Teglas et al., 2011). Further, a study with 20-month-olds corroborates findings of probabilistic abilities, particularly the capacity to use statistical information to infer human preference. After seeing someone remove five of the same object from a container, both 20-month-olds and preschoolers inferred that person's preference for that object (Kushnir, Xu, & Wellman, 2007).

As children get older, they can use probabilities to think about complicated concepts, such as emotions. For example, four-year-olds succeed on tasks asking them to predict emotional outcomes from stories with varying outcome probabilities (Lagattuta & Sayfan, 2011). By age five, children expect that emotional state (i.e., positive versus negative) exerts an impact on cognitive performance, specifically performance on an academic task (Amsterlaw, Lagattuta, & Meltzoff, 2009). These findings suggest that children can use information about outcome probabilities – from information provided by an experimenter and from their own experience – to inform their predictions about complicated concepts such as emotions.

However, despite findings that children have complex probabilistic reasoning skills by age five, prior literature suggests that preschoolers – aged three years to six years – understand causes of disease as determinate, i.e., exposure to germs necessarily makes one sick and there is no probability attached to that outcome (Kalish, 1998). One illustrative study presented five-year-olds with several stories about children being exposed to germs and disease agents, asking them two questions at the end of each story: "What do you think, will X get sick?" and "Do you know for sure or just maybe?" The first question asked for sickness predictions and the second served as a proxy for probabilistic thinking. These stories were accompanied by others with definite and variable outcomes unrelated to health, in which similar questions were asked. At age five, children were appropriately and accurately uncertain for stories with variable outcomes that were not about health, but they perceived causes of illness as determined, not uncertain (Kalish, 1998).

Given these conflicting findings, it is possible that probabilistic reasoning is domain-specific; perhaps the health domain is more complex than other domains previously tested with young children. After all, many of the studies conducted to assess probabilistic reasoning are rather contrived; children use statistical input to make predictions about a future action, e.g., whether a person will select a pink ball or a yellow ball from a bin in which there are four times as many pink balls as yellow balls (e.g., Denison et al., 2013). Such studies do not carry the complexity of illness transmission, which includes multiple actors who are often invisible and who engage in interactions over time in which any cause-and-effect relationship is necessarily delayed. Children may be able to use statistical input to consider relative probabilities in only relatively simplified scenarios and may not be able to extend such probabilistic reasoning to more complicated, realistic domains.

It is also possible, though, that young children’s difficulty with probabilistic reasoning in the health domain relates more to the nature of the problems about which researchers are asking them to reason. Scenarios like choosing a ball from a bin or identifying emotional reactions to certain situations incorporate aspects of *human agency*. In the types of health-related stories that have been posed to children by researchers such as Kalish (1998), humans were not agents; they were patients, inflicted with certain symptoms from germ (non-human, invisible) agents. Perhaps children can reason probabilistically when a human is the agent – when someone is selecting a ball or feeling an emotion – but cannot extend that kind of reasoning to a germ agent with a human patient. Based on his data, Kalish (1998) ultimately concludes that preschoolers do not extend their probabilistic reasoning skills to the health domain; he even suggests that this inversion of agency, where the human becomes the patient and loses agency, may explain this lack of extension. It is unclear, though, whether the lack of extension is rooted in the aforementioned inherent complexity of health and illness as concepts or rather can be attributed to the unique circumstance of combining non-human agents, i.e., germs, with human patients.

Preventative health behaviors introduce human agency into the traditional illness transmission mechanism (Figure 1). Choosing to wash one’s hands or clean an apple before eating it, for example, are active, agentic steps taken to prevent illness. By disrupting the traditional transmission model – from illness irritant to human – that children took to be definite in Kalish’s (1998) study, preventative health behaviors change the probabilities associated with illness irritant (i.e., germ) exposure.

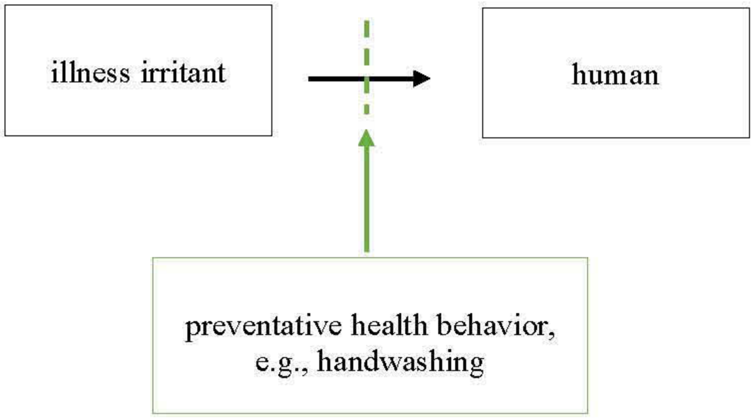


Figure 1. The illness transmission mechanism disrupted by preventative health behavior. Preventative health behavior disrupts illness transmission – a cause-and-effect relationship children take to be certain – and introduces an aspect of human agency.

Do preschoolers view preventative health behaviors as probabilistic or deterministic? If preschoolers interpret stories containing preventative health behaviors as deterministic to the same degree as stories without preventative health behaviors, one can conclude that the inherent complexity of the health domain inhibits probabilistic understanding. On the other hand, if preschoolers interpret preventative health behaviors as leading to probabilistic sickness outcomes (i.e., less deterministic in their outcome than stories without such behaviors) such responses might suggest the presence of health-related probabilistic reasoning in young children. In order to investigate these questions, the following pair of studies explored probabilistic reasoning in the health domain in both adults and young children.

In both studies, participants were presented with stories involving illness transmission and with characters either engaging or not engaging in preventative health behaviors. Participants predicted whether the main character in the story would get sick and gave certainty judgments ("for sure" or "just maybe") after predicting sickness. These certainty judgments were used to draw conclusions about the degree to which adults and young children think probabilistically about preventative health behaviors.

Study 1a: Probabilistic Reasoning in Adults

In this study, adults were presented with stories about characters engaging or not engaging in preventative health behaviors. After reading each story, participants answered two multiple-choice questions. The first question provided a sickness judgment: "What do you think, will X get sick?". The second question provided a certainty judgment: "Do you know for sure, or just maybe?". These two judgments constituted the primary dependent measures for the following study.

The purpose of the adult component of this study was two-fold. First, I aimed to explore the differences in certainty judgments for sickness-relevant stories between those involving the presence of preventative (e.g., washing one's hands or cleaning dirt from vegetables before eating them) and those not involving preventative health behaviors (e.g., sharing a straw with a sick friend or eating from a plate with bugs on it). Previous studies (e.g., Kalish, 1998) have explored children's probabilistic causality in the health domain using stories containing non-preventative health behaviors. This study augments existing literature by incorporating preventative health behaviors.

For the purposes of this study, presence or absence of preventative health behaviors is referred to as "agentic condition." Given the greater salience of human agency in preventative health behaviors than in non-preventative health behaviors, it was hypothesized that adults would be more probabilistic (less certain) in their predictions of sickness outcomes when exposed to stories containing preventative health behaviors. In other words, when exposed explicitly to a preventative health behavior (e.g., washing one's hands), adults would be less certain in their judgment of whether sickness would result or not. The study was run with manipulations of agentic condition both within-subjects (all participants read stories containing both preventative and non-preventative health behaviors) and between-subjects (participants read stories containing either all preventative health behaviors or all non-preventative health behaviors).

The second purpose of this study was to explore the role of what Kalish (1998) termed *potency* to present stories involving differing strengths of relationships between sickness agents and possible outcome. Kalish separated test items into either "strong" (e.g., hugging a sick child) or "weak" (e.g., waving to a sick child) causes of illness. Although the difference in potency did not influence children's predictions of sickness or their certainty of the outcome, adults did differ in both predictions of illness (expecting strong items to cause illness more frequently than weak items) and certainty of outcome (thinking of weak items as more certain). For example, it may well be that eating a sandwich previously discarded in a public garbage receptacle will likely generate more positive sickness predictions from perceivers ("yes, X will get sick"), along with less certainty with regard to sickness predictions, than eating a freshly-made sandwich. Accordingly, it was my hope that, based on adult responses, stories could be identified in which children had the best chance to be probabilistic in their responses. In short, the intention was to use these potency-dependent sickness examples and certainty judgments to inform the stories later used in Study 1b with young children.

To investigate the importance of potency, three types of stories were created: high likelihood of sickness (high-potency), medium likelihood of sickness (medium-potency), and low likelihood of sickness (low-potency). Examples of potency include spending time in a hospital (high-potency), in a clothing store (medium-potency), and alone at home (low-potency).

For both the within-subjects and between-subjects designs (which involved manipulations of agency), potency was always manipulated within-subjects.

I hypothesized that sickness judgments would mirror the potencies – such that high-potency stories would produce the most positive sickness judgments, followed by medium-potency, then followed by low-potency. Consistent with past research findings, I also hypothesized that certainty judgments would be greatest for low-potency stories. That is, when the likelihood of sickness was low (e.g., staying home alone and not interacting with any illness irritants), participants would provide negative sickness judgments (i.e., say "not sick") and would be fairly certain about their judgments.

With two independent variables – the presence of preventative health behavior (two levels: preventative and non-preventative health behaviors) and potency (three levels: high, medium, and low likelihoods of sickness) – the study employed a 2x3 design. For the within-subjects design, each participant read a different story associated with each of the six levels of the relevant variables (e.g., high-potency and agentic-preventative, low-potency and non-agentic-non-preventative, etc.). For the between-subjects design, only one agentic condition was presented across all potencies for any given participant.

Method

Participants. A total of 157 participants (61 female) were recruited using Mechanical Turk; 80 participants (30 female) were assigned to the within-subjects design and 77 participants (31 female) were assigned to the between-subjects design. The study was conducted using an online survey created with Qualtrics software licensed to Swarthmore College. Through the Mechanical Turk compensation program, participants were each paid $1.50 for their participation.

Materials. Five scenarios were created, each of which involved three different potencies (high, medium, and low), resulting in fifteen stories overall. Each participant was presented with each scenario in each of its three potencies. In the between-subjects version, participants saw either all fifteen stories with preventative health behaviors or all fifteen stories without preventative health behaviors. In the within-subjects version, participants were randomly assigned to receive either the preventative or non-preventative behavior agentic condition for each story. Every participant received a near-even balance of stories with and without preventative health behaviors (eight preventative and seven non-preventative, or vice-versa). Breakdown of scenarios, potencies, and agentic conditions can be seen in the hierarchical model in Figure 2.

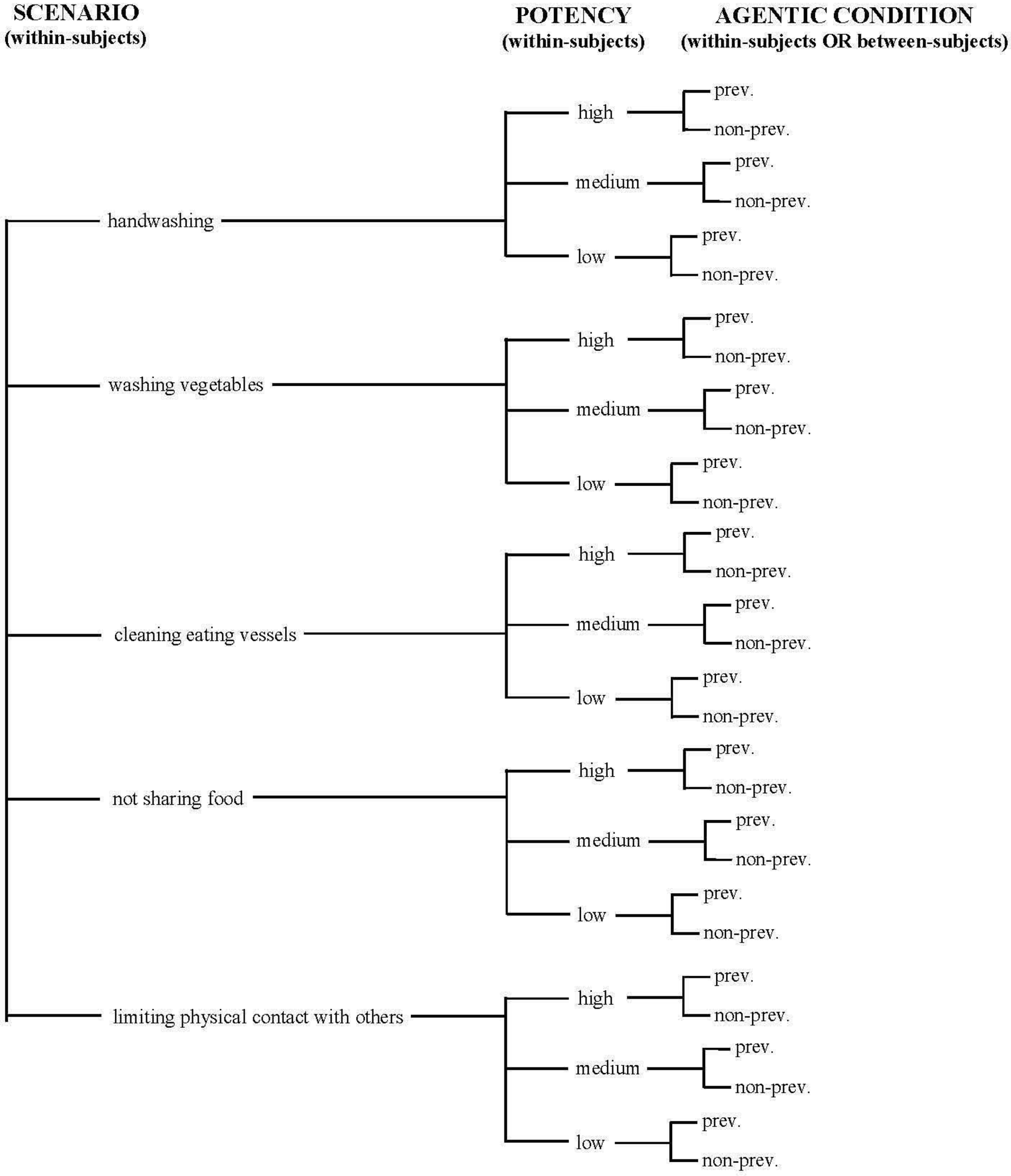


Figure 2. Study 1 design. Scenario and potency were varied within-subjects. Agentic condition (far right) was varied either within-subjects or between-subjects, depending on participants' assignment to one design or the other.

Additionally, participants completed six "catch trials" to ensure attention. Four of these test trials were written to expect definite certainty judgments whereas the remaining two were written to expect indefinite certainty, or "variable," judgments. Two of the definite and both of the variable catch trials were identical those used by Kalish (1998). The remaining definite catch trials were developed by the author. In total, participants were presented with 21 stories: 15 test trials plus six catch trials.

Procedure. Participants consented to participation in the study. They were given a brief synopsis of the experiment, asked to answer the questions intuitively, and told that the stories would eventually be presented to young children. Participants completed two practice trials with evaluative feedback (e.g., "That's correct – we would expect Kyle to come down to the ground because of gravity") before the study began.

Participants were presented with stories one at a time and asked two multiple-choice questions: "What do you think, will X get sick?" (choose either "yes, X will get sick" or "no, X will NOT get sick") and "Do you know for sure, or just maybe?" (choose either "for sure" or "just maybe"). As indicated above, each participant read twenty-one stories. Also as previously indicated, fifteen of these stories were derived from the five scenarios, with three potencies each. The remaining six were catch trials: two definite outcome and two variable outcome trials from Kalish's (1998) study and two additional definite outcome trials created by the author. The order of the stories was randomized.

Participants were then asked to complete four demographics items through the online survey: age, gender, race/ethnicity, and highest level of education obtained. Finally, they were thanked for their participation and received compensation through Amazon's Mechanical Turk.

Dependent measures. This study included two dependent measures: sickness judgments and certainty judgments. A positive sickness judgment was defined as the percentage of participants indicating that the actor in the story would get sick. This measure was calculated both for each story and, for each participant, for groups of stories based on agentic condition (i.e., preventative versus non-preventative health behavior) and potency (i.e., high, medium, or low).

Consistent with Kalish’s (1998) conceptualization, certainty judgments were considered a proxy for deterministic or probabilistic thinking, in which responses of “know for sure” were judged deterministic responses whereas “just maybe” were judged probabilistic responses. For purposes of the data analysis, a judgment of certainty was defined as a “know for sure” response.

Data coding, processing, and analysis. Percentage of sickness judgments and certainty judgments were calculated for each participant. Positive sickness judgments were coded as “1” and negative sickness judgments were coded as “0.” Deterministic (“for sure”) certainty judgments were coded as “1” and probabilistic (“just maybe”) certainty judgments were coded as “0.” Because both sickness and certainty judgments were binary, logistic mixed effects regressions were performed using the R statistical package. Logistic mixed effects regressions, rather than other methods, were chosen in order to maximize statistical power. All other statistical analyses – in particular, ANOVAs and *t*-tests – were performed using SPSS.

Results

All effects obtained through the use of the within-subjects version of the study were as strong or stronger than those in the between-subjects version; thus, for brevity, only the within-subjects version results are reported here. Neither sickness nor certainty judgments differed between the two within-subjects agentic condition survey versions (all *p*s > 0.2). Data from both within-subjects survey versions are thus combined in the following analyses.

Sickness judgments. The "potency" independent variable manipulation was expected to alter the relative likelihood of sickness associated with each scenario. Accordingly, sickness judgments provided by participants served as a form of manipulation check. If the manipulation of potency was successful, sickness judgments should indeed vary by potency, with high-potency stories (e.g., spending time in a hospital or sharing a drink with a sick friend) producing the greatest number of positive sickness judgments. Indeed, as predicted, sickness judgments differed by potency. High-potency stories (judged by participants to involve the protagonist getting "sick" 55% of the time) differed from both medium-potency (judged "sick" 42% of the time) and low-potency (judged "sick" 11% of the time) stories (both *p*s < 0.001).

To corroborate these findings, a logistic mixed effects regression was performed with fixed effects for agentic-preventative condition, potency, scenario, and survey version, as well as random effects for participant. As predicted, for sickness judgments, there was a significant effect of potency ($\beta = 4.41$, $p < 0.001$), with high-potency stories producing the greatest number of positive sickness predictions, followed by medium-potency, then low-potency stories.

In sum, given the effects of potency reported above, the potency manipulation was successful: high-potency stories that were designed to suggest the highest likelihood of sickness did indeed result in sickness judgments by participants that reflected this intention.

Certainty judgments. A logistic mixed effects regression was performed for certainty judgments, with fixed effects for agentic condition, potency, scenario, and survey version, and random effects for participant. This model revealed significant effects of agentic condition ($\beta = 3.10$, $p < 0.001$), as well as potency ($\beta = 1.81$, $p < 0.001$) on certainty judgments, indicating that certainty judgments differed based on both the presence (or absence) of preventative health behaviors and on the likelihood of sickness.

Looking specifically at the effect of agentic condition (the presence or absence of preventative health behaviors), adults were probabilistic in their responses for both agentic (containing preventative health behaviors) and non-agentic (not containing preventative health behaviors) stories, but they were more probabilistic in response to non-agentic stories ($p < 0.001$). Stories with preventative health behaviors (e.g., those involving hand washing) were judged more certain (41% certain) with regard to sickness prediction than stories without preventative health behaviors (26% certain). All certainty judgments differed from chance (those involving preventative health behaviors: $t(79) = 2.33$, $p = 0.02$; those not involving preventative health behaviors: $t(79) = 7.84$, $p < 0.001$). In sum, adults were probabilistic in their responses overall and were particularly probabilistic for stories that did not contain preventative health behaviors.

Certainty judgments also differed based on potency. Participants were probabilistic (less certain) for high- and medium-potency stories, answering "for sure" significantly less frequently than chance (high: $t(79) = 6.26$, $p < 0.001$; medium: $t(79) = 6.15$, $p < 0.001$). Participants did not differ significantly from chance when they provided certainty judgments for low-potency stories, $t(79) = 1.56$, $p = 0.12$, indicating that participants were neither deterministic nor probabilistic for low-potency stories. Certainty judgments did not differ by potency between high- and medium-potency stories (28% certain in both potencies; $p = 1.00$). Certainty judgments were significantly lower, though, in both high- and medium-potency than in low-potency stories (44% certain in low-potency stories; both *p*s < 0.001).

When the analysis was run to allow for an interaction between agentic condition and potency, agentic condition was no longer significant ($\beta = 1.39$, $p = 0.44$) while there was a significant interaction between agentic condition and potency ($\beta = 1.48$, $p = 0.04$). The effect of potency remained significant ($\beta = 3.24$, $p < 0.001$). These findings suggest that the presence or absence of preventative health behaviors did not independently influence probabilistic thinking; it acted alongside potency, i.e., likelihood of sickness, to influence such thinking. Potency, nonetheless, did independently influence probabilistic thinking, with a lower likelihood of sickness seeming to facilitate reduced probabilistic thinking.

In terms of relative differences, participants were least certain (most probabilistic) for medium-potency stories without preventative health behaviors (19% certain) and most certain (least probabilistic) for low-potency stories with preventative health behaviors (55% certain). Table 1 presents mean percentages of certainty judgments across all types of stories.

Table 1

Mean Percentage of Certainty Judgments, by Potency and Agentic Condition

Agentic condition	Potency			
	high	med	low	overall
Contains preventative health behavior	32%	37%	55%	41%
Does not contain preventative health behavior	24%	19%	34%	26%
Overall	28%	28%	44%	34%

Note: Certainty judgments are defined as "know for sure" responses.

Discussion

As stated previously, the purpose of the study involving adult responses was twofold. First, I aimed to establish a comparison baseline with respect to agentic condition for use with the children's version of the study. Contrary to what was expected, adults were more probabilistic (less certain) about sickness judgments for stories without preventative health behaviors (e.g., not washing hands) than for stories with preventative health behaviors (e.g., washing hands). This finding may have been obtained because adults understand preventative behaviors to disrupt the infection pathway to such an extent that those behaviors ultimately eliminate the likelihood of infection. For example, perhaps handwashing is such a strong preventative behavior that it eliminates most risk of illness. More research is needed to understand the cognitive mechanism influencing adults' probability judgments in the health domain.

The second purpose of the adult version of the study was to identify the scenarios in which children had the best chance to display probabilistic responses. Based on the original hypothesis regarding the effects of increased presence and salience of human agency, it was expected that the presence of preventative health behaviors would encourage probabilistic thinking. Accordingly, the findings of this study were used to identify stories containing preventative health behaviors with a potency that resulted in the lowest percentage of certainty judgments. As the data indicated, the stories with the lowest percentage of certainty judgments when preventative health behaviors were included were the high-potency stories. Thus, high-potency stories were selected for use in the subsequent study – Study 1b – involving the responses of children.

Study 1b: Probabilistic Reasoning in Children

In this study, children were presented with stories about other children engaging or not engaging in preventative health behaviors, following the same procedure as for adults in Study 1a. All stories had high likelihoods of sickness, as indicated by the largely positive sickness judgments (more "sick" responses) provided by adult respondents to the relevant scenarios in Study 1a. High-potency stories (i.e., stories with high likelihoods of sickness, such as a character spending time in a hospital or obtaining vegetables covered in dirt) were chosen because, when preventative behaviors were included, adults were maximally probabilistic in their responses for those stories.

This study investigated whether children are probabilistic or deterministic when confronted with stories that contain preventative health behaviors, especially as compared to their predicted probabilistic reasoning in response to stories that do not contain such behaviors. Agentic condition – the presence or absence of preventative health behavior in any given story – was the only independent variable for the children's version of the study. Agentic condition was manipulated in a within-subjects fashion, such that each child heard two preventative and two non-preventative stories. That is, each child heard four stories, two of which contained preventative health behavior (e.g., hand washing) and two of which did not contain such behavior (e.g., sharing a straw). Just as in the adult version, each story was followed by two questions: "What do you think, will X get sick?" and "Do you know for sure, or just maybe?". These questions constituted the two dependent measures: sickness judgments and certainty judgments, respectively.

Method

Participants. A total of 21 children (9 female) between 4 years, 3 months and 5 years, 9 months (mean age = 4 years, 11 months) were recruited from local preschools. Four children (19%) gave the same certainty answer for each question, i.e., answered "just maybe" to every question, including the catch trials. These children were excluded from further analyses.

Parents or guardians of all children gave written consent before the experiment was conducted. Children were given a book or a t-shirt after completing the experiment.

Materials. Four of the five original scenarios from the adult condition were used. The fifth test scenario from the adult version was omitted because it did not include clear-cut differences in human agency. All stories contained the high-potency format because, as mentioned above, based on the results of Study 1a, these stories were predicted to afford children the best chance to be probabilistic. Each child heard six stories – four test trials, one definite catch trial, and one variable catch trial. The catch trials were the same for all subjects and were taken from catch trials used in the adult version, originally employed by Kalish (1998).

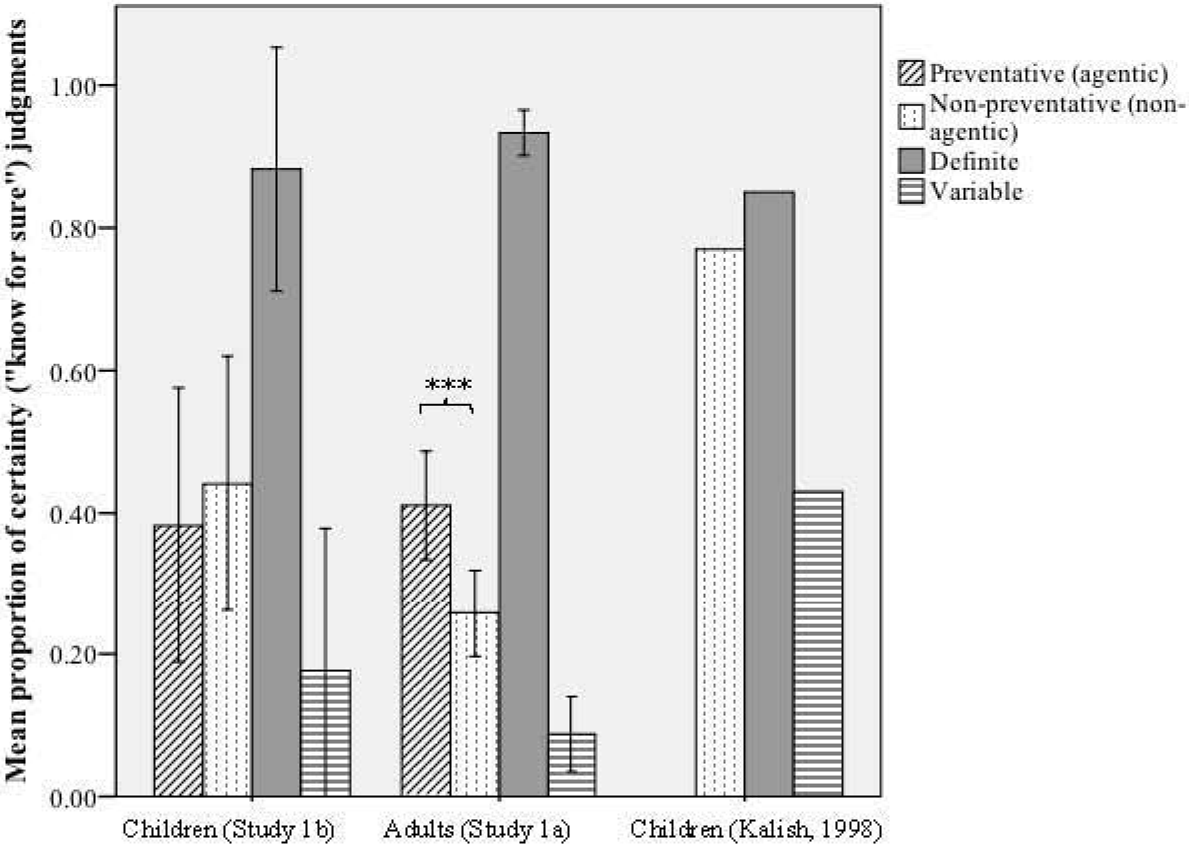
Procedure. The same procedure was used with children as with adults, except that the study was conducted in person as opposed to online, and the stories were read aloud to participants. After each new piece of information in the story, the experimenter asked children to repeat the information to ensure attention and comprehension. Corrective feedback was given to facilitate proper understanding of the stories (e.g., "Where did Joseph spend the day? That's right, he spent the day at the hospital."). The order of trials did not differ between participants.

Dependent measures. The dependent measures were the same as in Study 1a: sickness judgments and certainty judgments. I operationalized responses in the same way for both studies, i.e., certainty judgments were operationalized using "know for sure" and "just maybe" responses, as in Kalish (1998).

Data coding, processing, and analysis. The same coding scheme and analytical methods were used as in Study 1a. Logistical mixed effects regressions were performed using the R statistical package because of the binary nature of both dependent measures. All other calculations and dependent variable aggregations, e.g., means, were obtained using SPSS.

Results

Children provided "for sure" responses 41% of the time overall, far lower than the 78% certainty reported in Kalish (1998). They were certain 38% of the time for preventative and 44% certain for non-preventative scenarios. Certainty judgments from my studies with children and adults and from Kalish's (1998) study with children are shown in Figure 3.



Error Bars: 95% CI

Figure 3. Mean percentage of certainty judgments ("know for sure") for children and adults from Studies 1a and 1b and from children from Kalish (1998). (*** indicates significance at $p < 0.001$).

Children performed at chance to the extent that all logistic mixed effects regressions for both sickness and certainty judgments did not converge because of singular fit. In other words, the logistic model used in Study 1a with adults could not be used here because children's responses for both sickness and certainty judgments could not be reliably predicted using such a model. Thus, mean percentages for each trial are reported and variability between scenarios are discussed below, but formal statistical tests are not included for every relevant comparison.

Sickness judgments. Positive sickness judgments (answering "sick") were far higher for stories without preventative health behaviors (e.g., no handwashing) than for stories with preventative health behaviors (e.g., handwashing). These findings are consistent with what was predicted, especially because children heard stories, drawn from the adult version of the study, that featured a high likelihood of sickness. Sickness judgments for individual trials are shown in Table 2.

Table 2

Mean percentage of positive sickness judgments

	Scenario				
Agentic condition	Handwashing	Washing vegetables	Cleaning eating vessels	Not sharing food	Overall
Contains preventative health behavior	14%	0%	43%	50%	26%
Does not contain preventative health behavior	100%	100%	90%	100%	97%

Note: Positive sickness judgments are defined as "sick" responses.

A logistic regression analysis revealed a slight but statistically nonsignificant effect of age, with younger participants trivially more likely to answer "sick" than older participants ($\beta = 1.07$, $p = 0.14$). There was no effect of gender on sickness judgments ($\beta = 1.20$, $p = 0.72$).

Certainty judgments. There was great variability in certainty judgments both between and within scenarios. Children were most probabilistic in their responses for handwashing stories, both those that did and did not include the preventative handwashing behavior. Mean percentages of certainty judgments for each trial are shown in Table 3.

Table 3

Mean percentage of "know for sure" responses

	Scenario				
Agentic condition	Handwashing	Washing vegetables	Cleaning eating vessels	Not sharing food	Overall
Contains preventative health behavior	0%	70%	14%	50%	38%
Does not contain preventative health behavior	10%	71%	60%	43%	44%

Note: Certainty judgments are defined as "know for sure" responses.

Two children (12%) provided probabilistic responses for all test trials. No children were deterministic for all trials, but four children (24%) were deterministic for three of four test trials. All of the children who provided deterministic responses to the majority of trials happened to be in the condition in which the first test trial did not contain preventative health behavior (i.e., the first test trial featured a character who did not wash his hands).

An additional logistic regression analysis showed no significant effects of either gender ($\beta = 1.13$, $p = 0.81$) or age ($\beta = 1.04$, $p = 0.34$) on certainty judgments.

Discussion

Children's certainty judgments did not appear to differ between stories that did and did not include preventative behaviors (e.g., included handwashing versus did not include handwashing). However, compared to the 78% contagion reported for stories about illness obtained by Kalish (1998), children's overall certainty judgments in this study were far lower – 41% overall. That is, in contrast to earlier findings, children in this study appeared to employ probabilistic reasoning. There are several possible explanations for this discrepancy.

First, recall that Kalish (1998) presented children with stories in which characters never engaged in any preventative health behaviors; stories differed based on what was causing the illness, but the role of the character in disrupting that causal mechanism was always a passive one. This study, on the other hand, featured both characters who took active roles by engaging in preventative health behaviors such as handwashing and characters who were passive and did not engage in such behaviors. It is possible that natural comparison between preventative and non-preventative stories may facilitate children's ability to reason probabilistically. Simple exposure to variability in human behaviors – in this study, children heard stories where characters both engaged in and did not engage in preventative health behaviors – may highlight unpredictability and consequently encourage probabilistic thinking.

Second, it is possible that the situational context has changed since the initial studies were conducted in the 1990s; over the past two decades, the expansion of the Internet, the use of sophisticated hand-held technology, and the growth of new forms of media have vastly diversified and commodified the means through which children can learn about health and illness. Children's TV programs, storybooks, and educational materials may be increasingly geared toward preventative behaviors,[1] and familiarity with these concepts may improve probabilistic reasoning abilities.

Probabilistic reasoning in different scenarios. According to the present study, children are more likely to be probabilistic in their reasoning – that is, much less certain – than they were two decades ago (Kalish, 1998), but it is also critical to note the variability in the certainty judgments they provided in this investigation. On two of the test trials – the stories about consuming vegetables covered in dirt and eating off of an unwashed plate with bugs on it – children's certainty judgments were much closer to the ones offered by subjects in Kalish's (1998) study. Children, like adults, draw directly from their experiences; more familiar tropes of illness and health behaviors may lend themselves better to probabilistic thinking. In the present study, almost all children were probabilistic when presented with a character who either washed or did not wash his hands – a health-relevant behavior that is extremely familiar to most children, as handwashing is one of the most common preventative behaviors discussed with young children. Enter any preschool and one will likely find many colorful posters with detailed

---

[1] For example, the children's animated TV show series *The Magic School Bus* often features excursions into human bodies to understand the immune response. The remake of the original series, *The Magic School Bus Rides Again*, features an episode about illness transmission, the difference between allergies and traditional illness, and the importance of covering one's mouth when sneezing (Bloom & Weston, 2017).

instructions and pictures about when and how one should wash one's hands. Perhaps it is easier to be probabilistic when the situation is familiar; situations regarding food preparation may be more unfamiliar, whereas situations involving handwashing (or lack thereof) may be more familiar to young children.

The greatest difference in results between the two agentic conditions, i.e., the inclusion or exclusion of preventative health behaviors, was found in relation to the scenario involving a plate with bugs on it and then either washing (14% certain) or not washing (60% certain) the plate before eating off of it. In this case, it is possible that children's category for "bugs" is different than their categorization of germs. Indeed, prior literature suggests that preschoolers assign biological and psychological properties to ants and not to germs (Solomon & Cassimatis, 1999, Study 4). If we subscribe to the original hypothesis that the presence of human agency facilitates probabilistic reasoning, perhaps the presence of bugs which, one could argue, possess some humanlike qualities, adds an additional probabilistic boost to the scenario when the human character engages in preventative behavior. When the human character does not engage in preventative behavior, the human agency is not highlighted, and the humanlike properties of the bugs may not be as salient. Of course, this explanation is speculative and further investigation into the ways in which children understand this particular scenario is needed.

Comparing probabilistic reasoning in children and adults. In order to understand the development of probabilistic understanding in the health domain, it is valuable to compare adults and children. Interestingly, adults and children exhibited about the same level of certainty – indeed, children may have even been a little more probabilistic than adults – for stories that included preventative behaviors: adults were 41% certain and children were 38% certain of their sickness judgments. Adults and children differed in their level of certainty for stories that did not include preventative behaviors, in which adults were only 26% certain and children were 44% certain. Here, although the difference between adults and children obtained was not nearly as large as that reported as in Kalish's (1998) study, one can observe the same certainty differences between adults and children for stories without preventative health behaviors as in the health-related stories Kalish (1998) presented. Comparing these non-preventative scenarios with ones that involve preventative behaviors may facilitate increased probabilistic reasoning for children, as the relevant gap between adults and children decreased in this study. More strategies may be needed to encourage preschoolers to understand non-preventative health behaviors in the same way that adults do.

Particularly when thinking about learning, we evaluated children's performance in this study and in prior studies by comparing them to adults, finding differences in type and depth of cognitive reasoning. But what if their increased certainty in the non-preventative realm as compared to that of adults is actually helpful? Probabilistic reasoning may not serve one best when one thinks about health outcomes; thinking deterministically may encourage greater adoption of preventative health behaviors.

For example, assume that the causal mechanism surrounding handwashing and lack of illness is in place. If someone takes such a mechanism to be determined, that person will likely wash his or her hands more frequently than someone who thinks about handwashing as a means to an undetermined end. With young children who may not be able to perform complicated risk assessment calculations, thinking about scenarios that do not include preventative behaviors as necessarily making one sick may actually encourage engagement in a preventative health behavior. In order to think about engagement in these behaviors or not, though, one must have an intact understanding of the underlying causal mechanism of illness transmission and how handwashing disrupts that mechanism. Part II of this thesis explores causal understanding and learning in preschoolers.

Part II: Causal Understanding and Its Development

Accurate and holistic understanding of health and illness necessitates causal reasoning and understanding. Adults have strong causal reasoning skills; they have a firm grasp on causal reasoning in physical, emotional, and biological domains (e.g., Cheng, 1997; White, 1988). Children also possess complex causal reasoning skills; studies indicate that even six-month-olds can display preliminary causal understanding (Leslie & Keeble, 1987).

Prior literature suggests that five-year-olds have a piecewise understanding of illness, its causes, and its contagion. Specifically, preschoolers – three- to six-year-olds – connect contagion to symptoms, rather than to causes of disease; to a five-year-old, a cough is always contagious, whether that cough is caused by inhaling germs or by smoking a cigarette (Solomon & Cassimatis, 1999). One possible explanation for this outcome is that preschool-age children are unable to reason causally. However, children learn many relational concepts before the age of 5 – e.g., physical forces, familial relationships, emotional reactions, etc. – and it is clear that even young preschoolers possess these types of relational understandings (e.g., Buchanan & Sobel, 2011; Göksun, George, Hirsh-Pasek, & Golinkoff, 2013; Solomon, Johnson, Zaitchik, & Carey, 1996). For example, in one study, preschoolers were asked to demonstrate causal understanding of force dynamics using a board game. Although only older children – over five years old – were able to integrate opposite forces, three-year-olds demonstrated basic understanding of forces and could correctly predict the trajectory of a ball (Göksun et al., 2013). Preschoolers can also think about less salient, "invisible" causes; one study found that four-year-olds understand the difference in origin between natural (e.g., trees, oceans, and flowers) and artificial (e.g., cups, hammers, and shoes) objects (Gelman & Kremer, 1991). Preschoolers are clearly able to acquire sophisticated understanding of causal concepts across multiple domains, indicating that their limited ability to think causally in the health domain likely cannot be explained by a domain-general deficit in causal reasoning.

An alternative explanation for children's limited ability to think causally about health concepts may lie in the asymmetry of causal concepts and the intersection of that asymmetry with the health domain. Causal knowledge is direction-dependent; even adults find it much easier to reason causally from cause to effect than from effect to cause (Bright & Feeney, 2014; Kahneman & Tversky, 1973). Because effects of illness – i.e., symptoms – are much more salient and visible than causes of illness – e.g., germs, poor lifestyle habits, etc. – it is possible that this causal connection is particularly difficult to make.

Comparison

Given the complexity in causal asymmetry, it is useful to consider strategies that facilitate relational understanding. Causal understanding is a relational concept; in order to understand cause and effect, one must be able to relate the cause to the effect. For example, a car moving toward a stationary car will hit the stationary car and the stationary car will move. The cause of the movement of the stationary car is the collision with the moving car and the effect is the movement of the stationary car. In order to understand this physics concept, one must relate the movement of the two cars before the collision and after the collision. Multiple educational strategies have been proposed to support children to learn relational concepts; one such strategy is comparison.

Comparison has been shown to improve understanding of relational concepts. In particular, facilitating direct comparisons between various scenarios encourages correct categorization of objects and more accurate understanding of complicated causal relationships. Proponents of a strategy called *analogical bootstrapping* propose that comparison of two partially-known events facilitates category learning, including relational and causal categories (Kurtz, Miao, & Gentner, 2001). For example, in one study, adults were presented with two scenarios (pancakes in a frying pan and a coffee cup with an ice cube) and asked to participate in, what has been termed in the field of developmental psychology, an "alignment activity," which highlighted the similarities between the scenarios. Participants who successfully aligned the novel scenarios acquired a category of rules – heat transfer principles – that they could then apply to novel scenarios (Kurtz et al., 2001). Further research indicates that the alignment that occurs in analogical bootstrapping facilitates greater understanding and application of relational categories (Christie & Gentner, 2010; Kurtz, Boukrina, & Gentner, 2013). Learning by comparison occurs most successfully when the comparison is explicit and items are compared side-by-side (Mason, 2004).

All comparisons are not equal; similarity between the items being compared can dictate the efficacy of learning. The role of similarity in comparison is most frequently studied in category-based induction, where learners construct categories using various pieces of evidence. One is more likely to acquire (and apply) a novel category when the stimuli are more similar (Heit, 2000; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990). For example, given the information that dolphins eat fish, one is more likely to expect that whales eat fish than yaks eat fish, because dolphins and whales are more similar than dolphins and yaks. When abstracting a rule, though, it makes more sense to use items that are less similar in order to align properties and create a new category. Consistent with this logic, adults choose diverse, dissimilar evidence to learn new categories (Heit, 2000). Prior studies indicate that five-year-olds also select what the researchers designated low-similarity evidence over high-similarity evidence when making inferences about human preferences (Noyes & Christie, 2016). It is possible, therefore, that low-similarity comparisons may also facilitate more accurate understanding of disease causation and contagion.

The primary goal of this study was to explore potential educational strategies to support young children in differentiating between the contagion of germ- and event-caused symptoms. Informed by Solomon and Cassimatis' (1999) findings, children were presented with multi-part stories and asked them to complete the final panel of the story, after undergoing one of two learning conditions. Two central questions were investigated: First, does comparison facilitate greater understanding of disease causation and contagion, as measured by the ability to differentiate between the contagion of germ- and event-caused symptoms? Second, do different types of comparison – namely, comparison between similar or different pieces of evidence – support this understanding in different ways?

Study 2: Causal Reasoning in Children

In the present study, I examined the efficacy of comparison in helping preschoolers differentiate between germ (e.g., ingesting bacteria) and event (e.g., eating too much candy) causes of illness in predicting contagion. I did so by developing a novel comparison learning task. Participants were randomly assigned to one of two between-subject conditions: High-Similarity or Low-Similarity. During the learning task, participants compared two stories – one with a germ-cause and one with an event-cause – that differed to varying degrees based on the assigned learning condition. Following each learning phase, participants completed test phase tasks, in which they were asked to predict whether a child not exposed to the original illness irritant would get sick. This prediction task was accomplished by asking children to complete a pictorial sequence of illness transmission, i.e., by selecting in the final panel in a four-panel sequence. All participants were presented with the same eight stories (four learning trials and four test trials); the order of these stories and the combinations in which they were presented depended on the assigned learning condition. The primary dependent measure was participant choice of the "sick" or "not sick" card for each of the four test trials.

Visual representation of the model can be seen in Figure 4. Again, the purpose of the comparison learning task was to support preschoolers in differentiating between the two illness transmission mechanisms.

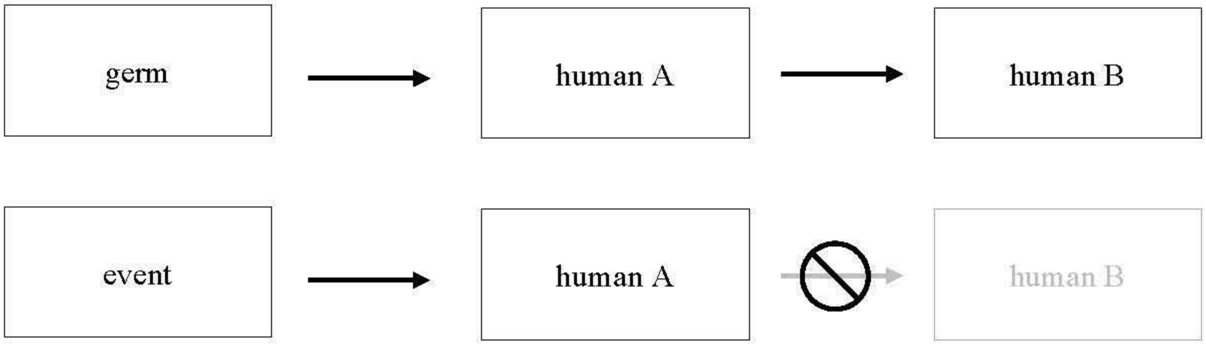


Figure 4. Illness transmission mechanisms for germ-causes (e.g., breathing in bacteria) and event-causes (e.g., smoking a cigarette) of illness. For germ-causes, illness is transmitted to human A, who can transmit illness to human B, i.e., the germ-cause is contagious. For event-causes, illness is transmitted to human A, who *cannot* transmit illness to human B, i.e., the event-cause is non-contagious.

Given the findings reported in prior studies (e.g., Solomon & Cassimatis, 1999), I expected that any form of comparison would facilitate improved causal reasoning and thus an enhanced ability to differentiate between the contagion of germ- and event-caused symptoms. I also expected that the two types of comparison – low-similarity and high-similarity – would facilitate causal reasoning to differing degrees.

Method

Participants. A total of 44 children (20 female) between 3 years, 1 month and 6 years, 6 months (mean age = 4 years, 9 months) were recruited from local day camps, preschools, and existing participant rolls from the Swarthmore Cognition and Development Lab. Parents or guardians of all children gave written consent before the experiment. Children were given a book or a t-shirt as a thank you after completing the experiment.

Materials. The study used eight stories, each of which contained four panels: cause, symptom, interaction, and interactive outcome (Figure 5). All stories were told with 3-inch by 3-inch laminated squares that featured black-and-white line drawings. Panels were laid out on colored felt with drawn-on squares that were the same size as the panels. The learning phase made use of red felt and the test phase, green felt. Both felt templates were placed on a table.

The eight stories included a germ and event cause for each of four symptoms: stomachache, cough, sneeze, and rash (see Appendix for all stories). Stomachache and cough stories were used for the learning phases, and sneeze and rash stories were used for the test phases. These symptoms were assigned as such because prior research suggests that preschoolers consider coughs and sneezes to be contagious symptoms and rashes and stomachaches to be non-contagious symptoms (Solomon & Cassimatis, 1999).

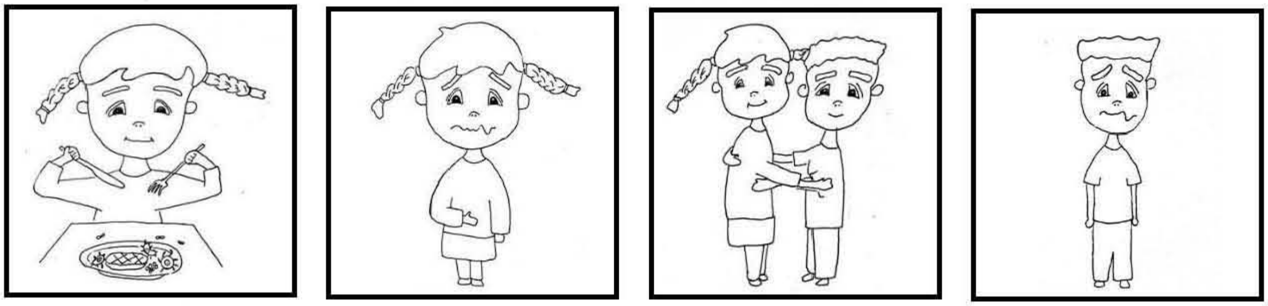


Figure 5. Four-panel story: girl eats food with germs (cause) – girl gets stomachache (symptom) – girl interacts with boy (interaction) – boy gets sick (interactive outcome).

Procedure. The study included two learning and two test phases (order: learning-test-learning-test), each of which contained two stories such that every participant heard and engaged with the same eight stories. Participants were randomly assigned to either the High-Similarity or Low-Similarity condition. In the High-Similarity condition, participants saw two stories portraying the same symptom in the learning phases. In the Low-Similarity condition, participants saw two stories portraying different symptoms in the learning phases. In both conditions, each learning phase included one germ-cause and one event-cause story. Learning phase setups are presented in Figure 6. Likewise, all test trial pairs contained one germ-cause and one event-cause, though those stories were not presented together and instead were presented in succession.

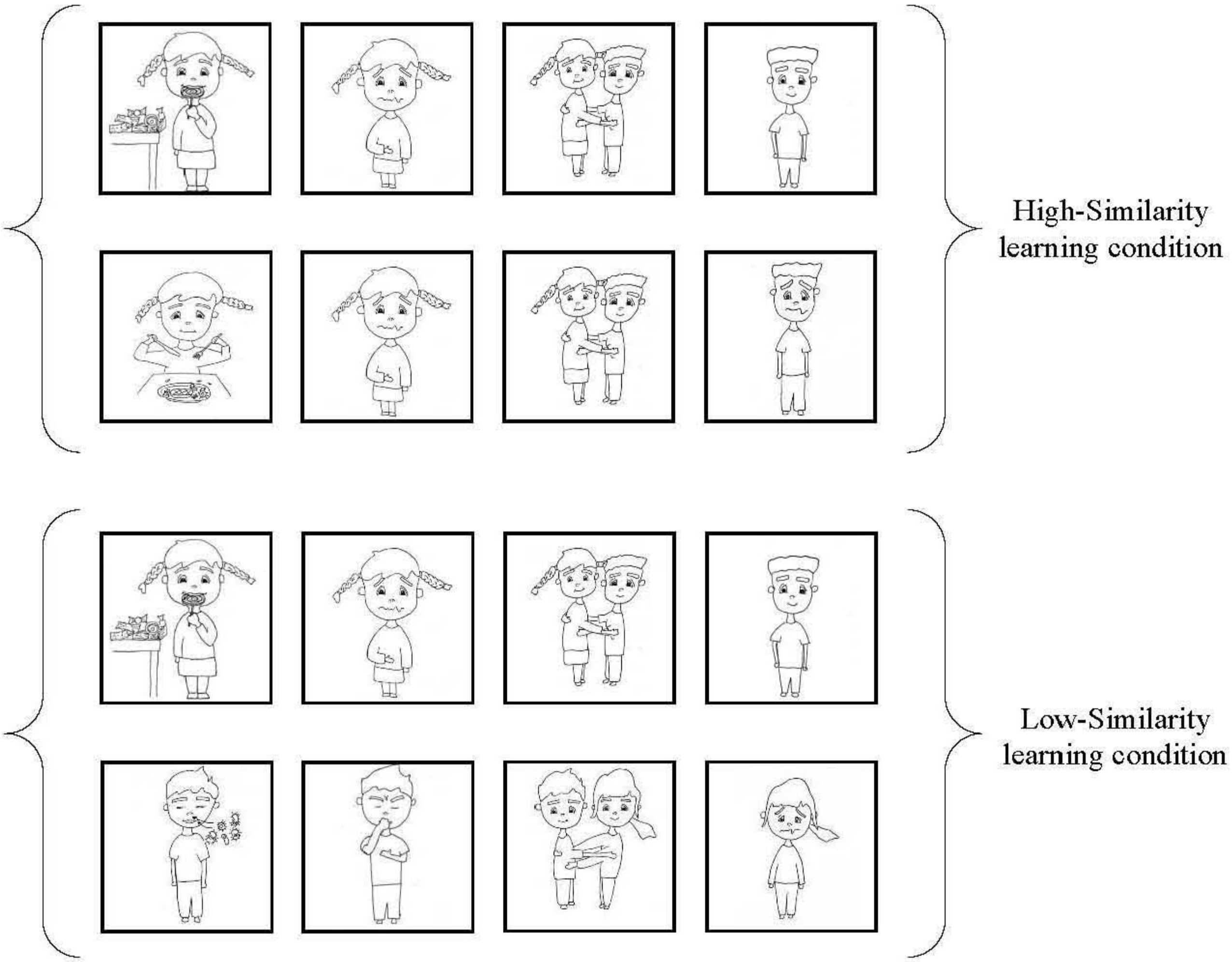


Figure 6. Examples of the two learning conditions – High-Similarity and Low-Similarity – as presented to children. In each example, the top row exemplifies an event cause (non-contagious) and the bottom row exemplifies a germ cause (contagious).

In the learning phases, the experimenter delivered the full story – i.e., using all four panels – with questions posed to the participant after each panel was presented (e.g., "Amy ate food with germs on it. What did Amy do?"), along with any corrective feedback as needed. After corrective feedback was given, the experimenter told the story again without employing the aforementioned questions. The second story in the learning phase was told in the same manner and the panels were laid out on the same surface, below the first story. After the experimenter delivered the second story, the experimenter asked the child to compare each of the panels of the two stories (e.g., "What happened to Amy here [pointing]? And what happened to her here [pointing to card below]?") and gave corrective feedback (e.g., "That's right – Amy did get a stomachache there") as needed.

In the test trials, the sequence of storytelling was the same as in the learning phase, but the experimenter did not provide the fourth (interactive outcome) panel. Instead, the child was given the option between "sick" and "not sick" panels and had to choose one card to place in the fourth spot. Unlike the design of the learning phase, the first test trial story was removed from the table before the second one was delivered.

The study design also contained one practice trial and one catch trial. The practice trial occurred at the very beginning of the experiment, before initiation of the task, and served to orient the participant to the act of choosing a card to complete a sequence of events. The catch trial occurred at the very end and served to ensure that the child understood the task. Both trials utilized age-appropriate cause-and-effect understanding in the development of the event sequence and were not related to health or illness (see Appendix).

Data coding, processing, and analysis. The primary dependent measure was participant selection of the "sick" or "not sick" card – referred to hereafter as "contagion judgments" – for each of the four test trials. Analyses were conducted for trials independently, for both contagion judgments and whether or not the response was correct (e.g., accurately understanding when the target child was likely to become sick). This secondary dependent measure is hereafter referred to as "accuracy." In addition, I also scored contingent understanding for each symptom (i.e., whether children correctly answered both germ- and event-cause trials for a given symptom). This measure is referred to as "differentiation."

I used logistic mixed effects regression – because contagion judgments were binary – to evaluate the effect of various factors on those contagion judgments. Pearson's correlations and non-parametric tests to compare distributions were run for aggregated data. Logistic regressions were run using the R statistical package and all other statistical analysis was performed in SPSS.

In terms of specific data coding, each trial was coded as correct or incorrect, and each child was given scores for the sneeze stories and the rash stories. Children that chose the correct card for both sneeze stories were given a score of 1 and likewise for the rash stories; in other words, children who differentiated correctly within a symptom were given a score of 1 for that symptom. Participants with only one of the two stories correct within a symptom (e.g., germ-rash correct but event-rash incorrect) were given a score of 0 along with participants who did not choose correctly for either of the stories. Holistic scores were out of 2, as the sum of the sneeze and rash differentiation scores.

I also coded and analyzed participant explanations for their choice using inductive qualitative content analysis. Quality of justification scores (which ranged from 0 to 4) were calculated for each participant for each trial based on consistency with contagion prediction, presence of sound causal logic, and articulation of correct health-related knowledge or understanding.

Results

Does comparison help? Children reported contagion at similar rates (58% for germs, 41% for events) as children in Solomon and Cassimatis' (1999) study (50% for germs, 38% for events). However, unlike Solomon and Cassimatis (1999), the addition of comparison in this study appears to have facilitated the children in this study's ability to differentiate between germ- and event-caused symptoms overall (Figure 7). I performed a logistic mixed effects regression with fixed effects for similarity condition, germ- or event-cause of the trial, rash/sneeze symptom of the trial, and quality of justification, as well as random effects for participant. With respect to contagion predictions, there was significant effect of cause type – germ or event – with higher reported contagion for germs compared to events ($\beta = 2.28$, $p = 0.02$). Reported contagion for both germs and events was given at chance (germs: $p = 0.06$; events: $p = 0.33$).

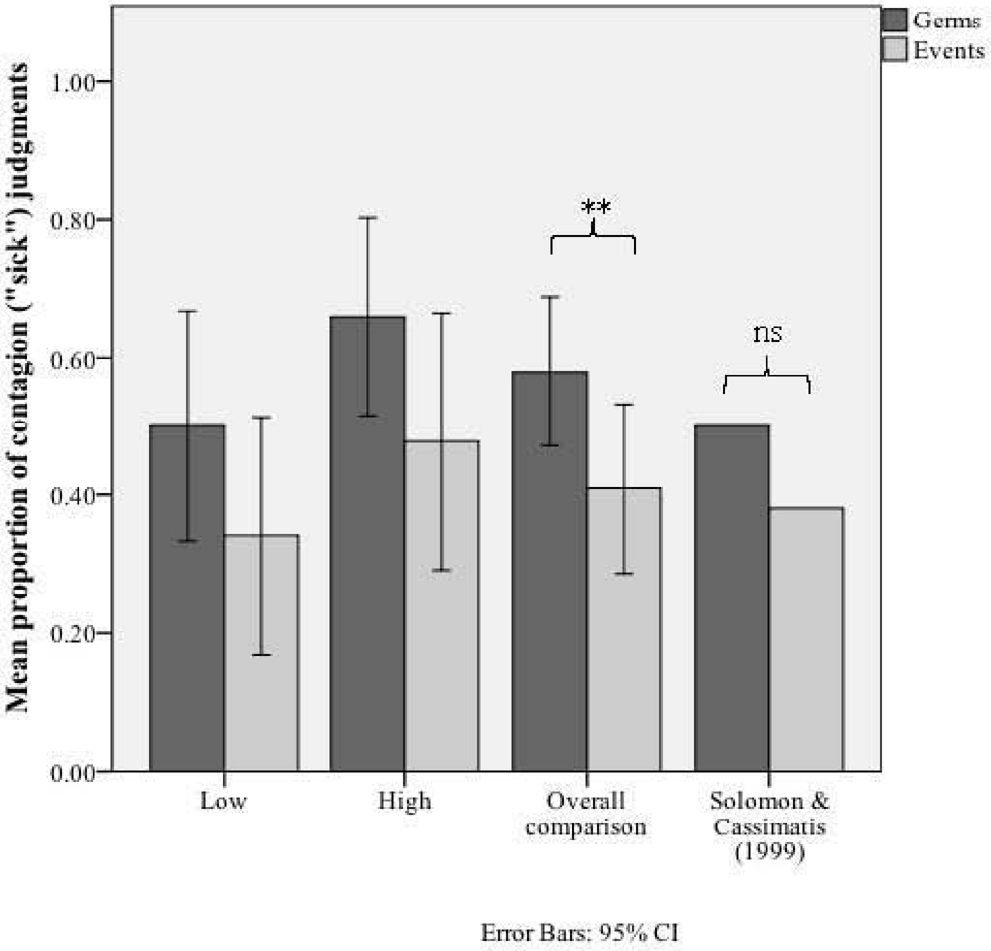


Figure 7. Mean percent predicted contagion, by germ- and event-cause. (** indicates significance at $p < 0.05$.)

For accuracy (e.g., choosing "sick" for a germ-cause test trial), a logistic mixed effects regression was run with fixed effects for similarity condition, germ- or event-cause of the trial, rash/sneeze symptom of the trial, and quality of justification, as well as random effects for participant. Only quality of justification had a significant effect on test trial accuracy ($\beta = 1.37$, $p = 0.02$; all other ps > 0.8). Accuracy on sneeze trials was not significantly correlated with accuracy on rash trials, $r(44) = 0.28$, $p = 0.07$.

In terms of differentiation by symptom (e.g., choosing "sick" for the germ-sneeze trial and "not sick" for the event-sneeze trial), children performed poorly, with only children succeeding on all four test trials only 29% of the time. Children differentiated germ- and event-caused sneezes correctly 27% of the time and germ- and event-caused rashes correctly 30% of the time. Differentiation between germ- and event-caused symptoms was below chance for each pair of test trials (sneeze: $p = 0.004$; rash: $p = 0.01$).

There were no gender differences for contagion judgments, accuracy, or differentiation outcomes (all ps > 0.18). Similarly, age did not predict any outcomes (all ps > 0.18).

What kind of comparison helps, if any? Although comparison learning helped children differentiate between germs and events in terms of contagion judgments, there was no main effect of similarity condition on any contagion judgments or performance accuracy. The logistic mixed effects regression model revealed, at best, a marginally significant effect of similarity condition for contagion judgments ($\beta = 2.12$, $p = 0.09$). Children in the High-Similarity condition reported contagion 66% of the time for germs and 48% of the time for events, compared to 50% and 34%, respectively, for children in the Low-Similarity condition. These differences are not significant (germs: $p = 0.53$; events: $p = 0.49$), and there was no interaction between germ/event cause and similarity condition for contagion judgments in the logistic mixed effects model (β = 1.07, $p = 0.92$).

Children in the High-Similarity condition (68% correct, on average) were more accurate for the germ-caused sneeze test trial than children in the Low-Similarity condition (41% correct, on average), a marginally significant difference, $t(42) = -1.85$, $p = 0.07$. All other test trials did not differ in accuracy between similarity conditions.

Correct differentiation by symptom and overall is shown in Figure 8. There were no differences in differentiation between similarity conditions.

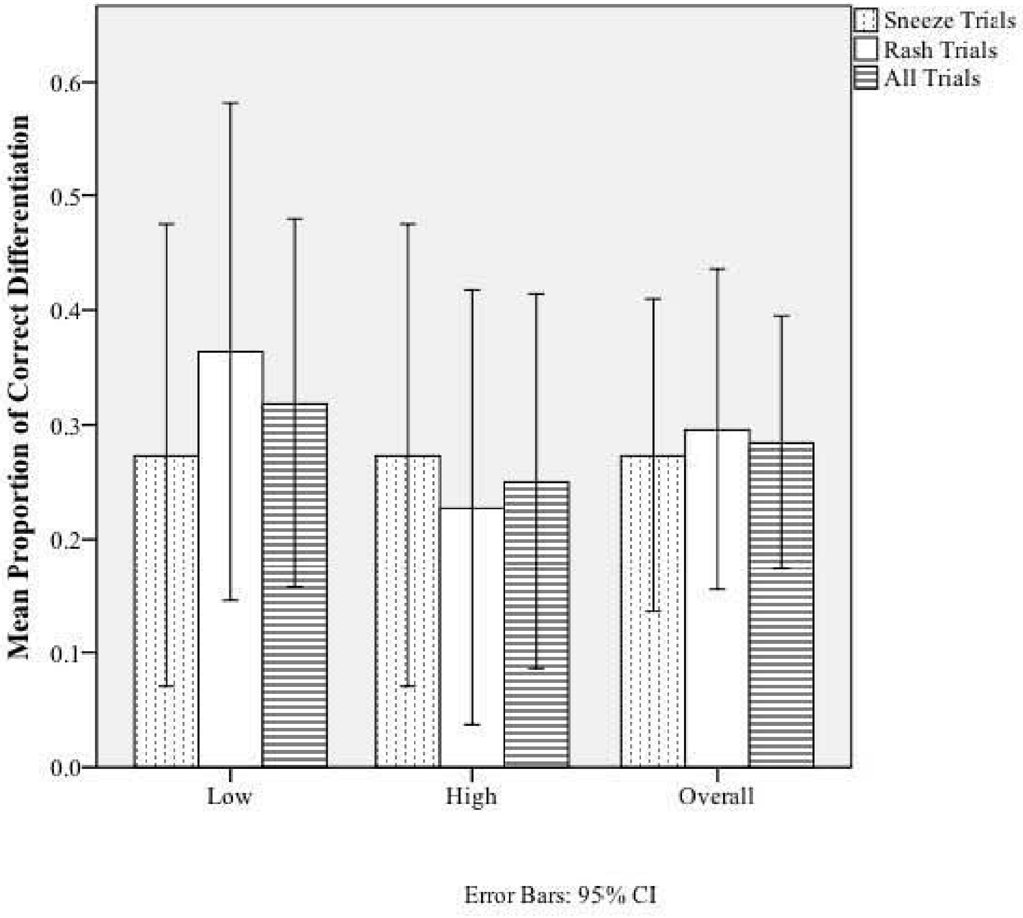


Figure 8. Mean proportion of correct differentiation by similarity condition and trial type (sneeze or rash) and overall.

Seven children (16%) answered “sick” for all test trials and 5 children (11%) answered “not sick” for all test trials. Children in the High-Similarity condition chose the “sick” card more frequently (57% of the time) than children in the Low-Similarity condition (42% of the time), but the difference was not significant ($p = 0.10$). Similarity condition did not predict answering all “sick,” $r(44) = 0.19$, $p = 0.23$, or all “not sick,” $r(44) = -0.07$, $p = 0.64$. Younger children were more likely to answer all “sick,” $r(44) = -0.30$, $p = 0.045$.

Qualitative analysis. Children’s responses to the open-ended question, “Why do you think X got / didn’t get sick?” were coded into six categories: 1) Irritant Nature – nature of the original irritant ($n = 45$; e.g., “it’s just flowers”); 2) Symptom Nature – nature of the symptom ($n = 28$; e.g., “because rashes do not switch to other skin”); 3) Interaction – reference to the interaction with the sick child ($n = 52$; e.g., “he got sick because he came over to play with her and she was sick”); 4) Non-Interaction – reference to not interacting with the original irritant ($n = 22$; e.g., “she didn’t get sick because she didn’t sit in the sun”); 5) Medical – invocation of medical or health terms ($n = 7$; e.g., “because it’s not contagious”); and 6) Fictionalization – invention of a story that explains the chosen outcome ($n = 15$; e.g., “when she was going to sneeze, she ran away and opened the door and closed the windows and then he didn't get sick”). Each explanation could be coded into multiple categories. See Table 4 for further examples of each coding category.

Table 4

Selected Quotes from Answers to the Question: "Why Do You Think X Got / Didn't Get Sick?"

Code category	Examples
Nature of original irritant	"Flowers don't make you sick." "It's germs and he was hugging her." "If you get a sun rash it can't go onto other people's skin."
Nature of symptom	"If she was sneezing a lot, [and] if she sneezed a lot when they were playing, then he would be sick." "A rash can affect other people."
Interaction with sick child	"She was sick and he hugged her." "She touched him." "She came over to play."
(Non-)interaction with the original irritant	"He didn't play with the toy that had germs on it." "She didn't sit in the sun where the sun is."

Invocation of medical or health terms	"It is contagious…contagious is when someone has something, and you get it from them." "He sat in the poison ivy and he was infected." "He's not allergic."
Fictionalization	"It was just germs and he might have washed his hands." "When she was going to sneeze, she ran away and opened the door and closed the windows and then he didn't get sick." "Tissues or her arms can block the sneezing from hitting him."

Note: Selected quotes are responses to the question, "Why do you think [character name] got / didn't get sick?"

Additional explanations included repeating the sickness judgment (*n* = 15; "she got sick because she got sick"), giving an answer unrelated to the story (*n* = 26), and "I don't know" statements (*n* = 22).

Some coding categories were used to explain both positive and negative predictions of illness. For example, children used Irritant Nature to describe contagion, in the case of germs or poison ivy, and non-contagion, in the case of sitting in flowers or spending time in the sun. These explanations were frequently accompanied by Interaction explanations (in the case of contagion) or Non-Interaction explanations (in the case of non-contagion).

Symptom Nature and Irritant Nature explanations were used in similar fashion; children would describe sneezes as inherently contagious and rashes as inherently non-contagious, a pattern also noted in Solomon and Cassimatis' (1999) study. However, Symptom Nature explanations were logical only when accompanied by Irritant Nature explanations, e.g., "If you get a sun rash it can't go onto other people's skin." Symptom Nature explanations frequently accompanied Fictionalization; children would embellish based on the symptom's presentation.

Medical and health terms were brought up to explain both types of predictions as well. Children who used terms like "contagious" or "infected" used those words in most, if not all, of their explanations.

Fictionalization answers were also used for both types of contagion predictions. Most fictional stories were told for the sneeze test trials. In these cases, children's stories often included preventative behaviors or materials, like handwashing or covering one's mouth during a sneeze, to explain why another child did not contract the sneeze.

Discussion

The purpose of Study 2 was to investigate the utility of comparison in facilitating children's understanding of contagion. Previous studies indicated that preschoolers do not differentiate germ- and event-caused symptoms in terms of their contagion (Solomon & Cassimatis, 1999). However, encouraging direct comparison between germ- and event-caused symptoms appears to facilitate differentiation between germs and events as disease agents. The two types of comparison (High- versus Low-Similarity) equally support children's causal reasoning about the differences in contagion for different causes of illness (i.e., germs versus events).

Although children in this study were differentiating between germs and events as the cause of sickness, their performance was still fairly poor. Complete understanding of disease causation and contagion would yield close to 100% contagion predictions for germs and 0%

contagion predictions for events. Children, both in this study and in previous literature, perform at chance for contagion judgments. Taken alone, this finding suggests that children's understanding of human-to-human disease contraction is incomplete and that children may not have a fully systematic way of thinking about disease causation in this context.

Children's explanations of their predictions indicate that they do think causally about disease contraction. Quality of justification was the only factor affecting test trial performance, suggesting that performance and ability to explain one's choice are closely connected. Interaction and Non-Interaction explanations require cognitive causal schemas. High proportions of this type of reasoning imply that some level of causal understanding in the health domain is indeed in place but may be underdeveloped and occasionally misapplied. Further, children often do understand the difference between germs and events as disease agents; in this study, many children invoked Irritant Nature descriptions to describe contagion and non-contagion. Although often used, causal mechanisms and understanding of the nature of irritants were not always used correctly (e.g., "It was germs and he didn't get sick because he didn't play with the germ toy"). Comparison may facilitate more accurate integration of these concepts, but further study is necessary.

Human-to-human disease contraction involves disease agents, but it also incorporates aspects of human agency that irritant-to-human disease contraction does not. Children's fictionalizations included preventative behaviors – classic and familiar examples of human disruption of irritant-to-human disease contraction – that involved characters taking steps to prevent the spread of illness. Although such behaviors were not presented or discussed at any point, children spontaneously incorporated activities involving either or both of the children in the story. Given that both germ and event contagion judgments in this study were at chance levels, perhaps children are probabilistic when it comes to predictions regarding disease contraction between individuals.

Past findings in the context of the present results. Solomon and Cassimatis (1999) conducted two studies asking children about the contagion of various germ- and event-caused symptoms. In the first study, children judged germ-caused symptoms to be contagious 84% of the time and event-caused symptoms to be contagious 72% of the time. In the second study, children judged germ-caused symptoms to be contagious 50% of the time and event-caused symptoms to be contagious 38% of the time. Children in both studies did not differentiate between germs and events in terms of contagion. The only difference in methodologies was that the second study included a line in each of the stories that explicitly connected the disease agent to the symptom (e.g., "A girl named Susan breathed in some pepper and pretty soon she got a runny nose" versus "A girl named Susan breathed in some pepper into her nose. Pretty soon the pepper in her nose made her have a runny nose"). Explicating the connection between the irritant and the symptom clarified the irritant-to-human disease contraction causal relationship. Perhaps the more explicit illustration of this relationship in the researchers' second study enabled children to think about the subsequent human-to-human disease contraction probabilistically. If that was indeed the case, two possible explanations could account for this improvement in understanding the probabilistic nature of disease contraction.

First, it is possible that explicating a relationship makes it more familiar. Unfamiliarity with a causal mechanism predicts unnatural and simplistic understanding of the causal relationship (Berzonsky, 1971). Because probabilistic thinking draws on a more complicated cognitive mechanism than deterministic thinking, more familiar concepts may be better suited to engaging probabilistic reasoning.

Second, making the non-human agent and human patient explicit may highlight the differences in the irritant-to-human and human-to-human disease contraction relationships. As discussed with respect to Study 1, the role of human agency and presence of human involvement in a causal relationship may make such a relationship more amenable to probabilistic thinking. Spontaneous incorporation of preventative behaviors in explanations for contagion judgments suggests that children are thinking about the irritant-to-human disease contraction mechanism to a degree sufficient to permit them to conceptualize disrupting it. As seen in Study 2, comparison may further highlight the nature – and variability – of the irritant-to-human disease contraction relationship and consequently facilitate probabilistic reasoning.

Granted, in all the relevant investigations, it is possible that children are paying more attention to the children in the story (rather than the symptoms or the illness irritants). If human-to-human relationships enable probabilistic reasoning and children are paying more attention to that relationship in the task, their answers could look probabilistic. However, if children are thinking differently about the human-to-human and irritant-to-human relationships, their forced-choice responses do little to illuminate these differences. Additional studies could seek to understand differences in children’s conceptualizations of these relationships by further unpacking illness transmission mechanisms.

It is also important to recognize that perfect accuracy on this task – in Study 2 and in Solomon and Cassimatis’ (1999) studies – asks children to be deterministic. The series of causal mechanisms through which the task asks children to reason engages multiple probabilities; compounding probabilities is notoriously difficult, even for adults. I explore this idea further in Part III, informed by the results of Study 1b.

Part III: General Discussion

This series of studies investigated children's conceptions of health, particularly in the domains of probabilistic and causal understanding. First, I examined children's probabilistic reasoning skills in the health domain, using findings from adults and prior literature as points of comparison. Second, I explored comparison as a learning strategy to help facilitate children's improved understanding of health-related causal concepts – i.e., disease causation and contagion. In order to better understand probabilistic causality and how it relates to the health domain – and these studies – it is useful to employ a model of illness transmission (see Figure 9).

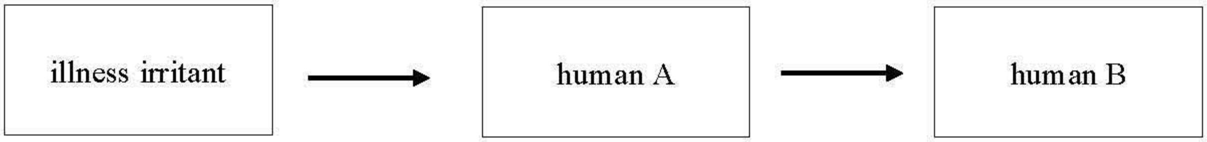


Figure 9. The three-part causal mechanism underlying illness transmission.

Figure 9 depicts the three-part causal mechanism underlying illness transmission; accurate and holistic understanding of illness requires such a mechanism. Studies 1a and 1b investigated probabilistic understanding with regard to the first arrow, that is, the transmission of illness from an irritant to a human. In his investigation of preschoolers' understanding of this relationship, Kalish (1998) suggested that preschoolers understand the arrow to connote certainty; namely, that exposure to an irritant would necessarily transmit the illness to a human. On the other hand, adults – in Study 1a presented here and in Kalish's (1998) study -- generally understand the arrow to indicate uncertainty; exposure to an irritant does not guarantee illness contraction. In this irritant-to-human relationship, the irritant is the agent and the human is the patient. Human agency is not salient and is only explicitly introduced when preventative behaviors, e.g., handwashing, are presented. When stories that incorporate preventative behaviors are presented with stories that do not include preventative behaviors, children’s overall judgment of certainty drops from 75% certain to 41% certain; they become much more probabilistic. Although in Study 1b children’s judgments of certainty did not differ with regard to stories with versus without preventative behaviors, their overall ability to think probabilistically about the irritant-to-human relationship improved when given stories of both types. It is possible that natural and automatic comparison between the two types of behaviors (i.e., preventative versus non-preventative) does indeed facilitate greater probabilistic thinking.

Study 2 investigated the full illness transmission mechanism (Fig. 9), from illness irritant to human A (primary) to human B (secondary). Prior studies suggest that children do not use the nature of the illness irritant to inform their understanding of the secondary human’s illness outcome (Solomon & Cassimatis, 1999). In Study 2 reported here, the experimenter presented children with two stories – one with a germ-caused symptom and one with an event-caused symptom – facilitating comparisons between them. This approach evidently supported preschoolers in their capacity both to abstract and apply a rule – a rule that enabled them to differentiate between the two types of illness irritants. Despite this improved differentiation compared to past studies, children still performed remarkably poorly on the task; reported contagion for all stories was around 50%. Their explanations corroborated previous studies indicating that only partial causal mechanisms are present in preschoolers (Solomon & Cassimatis, 1999); four-year-olds can reason from irritant to human and from human to human but cannot connect the two mechanisms. Because children can choose correct answers before they can articulate correct explanations or justifications for their choices, it may be that comparison facilitates a connection of these two-part causal mechanisms in a way that young children cannot yet articulate.

The findings from both Studies 1b and 2 – reporting children's conceptions of health-related probability and causality, respectively – suggest that preschoolers can think probabilistically in the health domain. Although they do not perform as adults do on all tasks, preschoolers possess at least partial causal mechanisms that they are able to articulate. Their low accuracy on the causality task may be a result of the complicated calculations required to compound probabilities. Because the illness transmission mechanism presented (Fig. 9) requires combining two probabilities (irritant-to-human and human-to-human), the calculation may be too complicated for young children, and their estimations of "probably yes" for germs and "probably no" for events may be off.

It is important to note that children in Study 2 demonstrated a desire to think about human agency and, perhaps, probability as they spontaneously incorporated preventative health behaviors into their explanations. As we saw in Study 1b, simply thinking about preventative health behaviors encourages probabilistic thinking. Even adults have trouble compounding probabilities; it is understandable that this calculation is difficult for preschoolers. More study is necessary in order to explore how children think about each component causal mechanism and how they associate various probabilities with each of those components.

Given these findings, I suggest two primary takeaways that may inform health-related educational practice. First, comparison may be a useful educational strategy in supporting children to understand illness, both probabilistically and causally. Second, children's existing knowledge and experience is very closely tied to their understanding of health-relevant questions and the reasoning skills they apply to this domain; acknowledgment and incorporation of children's existing experience may be useful in both exploring how kids think and how we may support them to learn. I expand on these claims below.

Comparison is helpful. Although only Study 2 explicitly examined the utility of comparison in supporting health-related understanding, investigating health-related probabilistic thinking in Study 1 indicated that comparison may also encourage probabilistic thinking. There is substantial literature on the efficacy of comparison for comprehending relational concepts such as causality (Gelman, Raman, & Gentner, 2009; Hoyos, Horton, & Gentner, 2015; Kurtz et al., 2013). Less is known about the utility of comparison in supporting probabilistic reasoning. This set of studies may contribute to a burgeoning literature regarding comparison and probabilistic reasoning.

In classroom discussions of health and illness, comparison may be a useful tactic when children are beginning to grasp concepts or have demonstrated partial understanding of the relevant domain. Directly comparing various health behaviors and their efficacy in preventing illness may encourage both more accurate causal understanding and more advanced probabilistic thinking. Comparison on the part of educators may come in the form of direct alignment – i.e., identifying similarities and differences between two example stories – or may involve simply juxtaposing scenarios involving different types of behaviors. Collaboration between psychologists and educators and innovation in incorporating research regarding how children think in specific fields may enable us to support children's understanding of health and illness more effectively.

Experience informs understanding. Children talk about what they know; their knowledge of health is primarily informed by their exposure to health concepts and materials. Quality of justification represented the only significant effect for accuracy in Study 2 – this finding tells us that kids' performance and ability to explain their reasoning are very closely connected. Perhaps we can infer that children who are more familiar with the health domain are better equipped to understand illness mechanisms and thus better equipped to explain their reasoning. After all, causal reasoning in other domains is demonstrably linked to familiarity with the domain (Berzonsky, 1971). If we subscribe to this explanation, causal reasoning in the health domain is domain-specific; causal ability in other areas does not necessarily facilitate causal understanding in the health domain. Further studies might compare causal reasoning in various domains in a within-subjects fashion to investigate its domain-specificity. Similarly, children's varying probabilistic reasoning skills – as seen in Study 1b – may be explained by differences in familiarity with particular health concepts, as alluded to above.

It is worth asking: Where does familiarity with the health domain originate? Certainly, children's understanding of the health domain is closely tied to that of their parents and families; mothers, for example, can predict their children's health and illness understanding remarkably well (Rubovits & Wolynn, 1999). But with expansion of the Internet, the use of sophisticated hand-held "smart" technology, and the growth of new forms of media, children are exposed to and learn about health concepts through many different channels. In addition, even traditional forms of media have incorporated new means of transmitting health knowledge. For example, television shows such as *The Magic School Bus* and movies such as *Osmosis Jones* teach children about how illnesses are transmitted, how the immune system responds to pathogens, and how preventative health behaviors might disrupt the illness transmission pathway (i.e., Fig. 9).

If greater familiarity with health and illness, even involving mere exposure to medical terminology, supports more advanced causal understanding, it need not be disruptive to the classroom to support such understanding in a preschool environment. Many preschool classrooms already encourage children to share their experiences and to talk about themselves; if this type of sharing is also incorporated into lessons about health, children may naturally become more familiar and competent with health and illness as concepts. Additionally, the use of media and various forms of technology in supporting both formal and informal health education needs to be explored further. These questions are especially important given that such means may in particular support children’s probabilistic causality in the health domain, where it has not previously been demonstrated.

Limitations

The present studies contribute to understanding the usefulness of comparison learning tasks, particularly in the health domain, and present some suggestions about how these findings may inform health-related educational practice. However, there are needed next steps for research on these topics. First, many of the claims made in this thesis are based on conclusions drawn in prior studies and use past findings as points of comparison. For example, results from the present studies suggest that comparison may improve children’s understanding of illness-relevant probabilistic causality; this argument is rooted in past researchers’ conclusions that preschoolers cannot reason about illness-relevant probabilistic causality. Although it is a common practice to use prior findings as points of comparison, this practice makes it difficult to conclusively ascribe differences in results between past and present studies to changes in methodologies or introduction of comparison tasks. In order to explore both the role of human agency and the efficacy of comparison as a learning strategy in the health domain, future studies may also seek to replicate past findings in addition to incorporating methodological changes. Second, child participants in the studies reported here were predominantly White and upper middle-class; further study would ideally include more diverse participants, allowing investigation of cultural and socioeconomic differences in how children conceptualize illness, especially given the relationship between understanding and lived experience. Third, as with many child development studies, sample sizes were small. Additional participants could illuminate patterns and differences that did not reach statistical significance in these studies.

Conclusion

To my knowledge, this is the first set of studies that seeks to investigate separately and then synthesize preschoolers’ reasoning concerning probability and causality in the health domain. In considering the results reported here, it is important to recognize the broad utility of comparison strategies in multiple domains while also realizing the unique nature and complexity of health and illness. Questions regarding preschoolers’ understanding of health are critically and necessarily tied to health outcomes because they presumably inform health behaviors. Given that conceptions of health and illness transmission in young childhood likely influence later understanding and behavior, questions investigated in the present thesis are particularly important as they appear to have implications for health-related educational strategies. Not only does this work suggest that children do have the capacity for probabilistic causal understanding in the health domain when supported by particular educational strategies, i.e., comparison, but it also indicates that further exploration in this realm may contribute to growing understanding in cognitive development, education, public health, and pediatrics.

Acknowledgements

Thank you to Stella Christie who, despite living out of the country for the entirety of this project, was still able to consult and advise me on this thesis, and without whom I could not have begun or carried out this project in the first place. Many thanks to Ann Renninger for her continued support, guidance, and feedback throughout the process and for encouraging a sense of humor and perspective in navigating the ups and downs of this research. I am very grateful to Andrew Ward for his patience and thoughtfulness in thinking about the written presentation of this work, especially in reading the document multiple times in very short timeframes. I am extremely thankful to Lisa Smulyan, for her compassion and mentorship in intellectual endeavors and in life, and for believing in me when I did not; to Molly Flaherty, for her insights, encouragement, and investment in me and this project; and to John Blanchar, for his invaluable statistical wisdom.

I am indebted to my wonderful parents, who were willing to talk through ideas with me at any hour and who have supported me to pursue anything and everything about which I care; to my brother, Krishna, for being my strongest and most enthusiastic cheerleader; and to my dear friends, who celebrated my greatest successes and coached me through what felt like insurmountable challenges. In particular, I am grateful to Amanda Izes and Shiloh Sumanthiran for reading pieces and drafts of this thesis; to Shira Samuels-Shragg for her constant ability to spread empowerment and love; to Chad Edwards-Kuhn for being a wonderful sounding board in all domains, providing wisdom far beyond my own, and always asking the most important questions; and to Rachel Isaacs-Falbel, Simona Dwass, Amy Shmoys, Jamie Starr, and Jason Guadalupe for their never-wavering encouragement and willingness to listen. Many thanks to the members of the Swarthmore Cognition & Development Lab for their creativity, patience, and never-failing ability to think outside the box. Thank you to Laura Chen for creating the stimuli for Study 2 and to MinJae Song for her help in creating the coding scheme in Study 2. Thank you, most of all, to the children who participated in these studies and their families for their willing participation – without them, this project would not and could not exist – and for all of the genuine joy they brought into my life over the past eighteen months.

References

Amsterlaw, J., Lagattuta, K. H., & Meltzoff, A. N. (2009). *Young Children’s Reasoning About the Effects of Emotional and Physiological States on Academic Performance. Child Development* (Vol. 80). Retrieved from http://ilabs.washington.edu/meltzoff/pdf/09AmsterlawMeltzoff_CD.pdf

Berzonsky, M. D. (1971). The Role of Familiarity in Children’s Explanations of Physical Causality. *Child Development*, *42*(3), 705. https://doi.org/10.2307/1127442

Bloom, G.A. (Writer), & Weston, R. (Director). (2017). Carlos Gets the Sneezes. [Television series episode]. In S. Stone (Producer), *The Magic School Bus Rides Again*. Netflix.

Bright, A. K., & Feeney, A. (2014). Causal knowledge and the development of inductive reasoning. https://doi.org/10.1016/j.jecp.2013.11.015

Buchanan, D. W., & Sobel, D. M. (2011). Mechanism-Based Causal Reasoning in Young Children. *Child Development*, *82*(6), 2053–2066. https://doi.org/10.1111/j.1467-8624.2011.01646.x

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367–405. https://doi.org/10.1037/0033-295X.104.2.367

Christie, S., & Gentner, D. (2010). Where Hypotheses Come From: Learning New Relations by Structural Alignment. *Journal of Cognition and Development*, *11*(3), 356–373. https://doi.org/10.1080/15248371003700015

Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology*, *49*(2), 243–249. https://doi.org/10.1037/a0028278

Evans, J. S. B. ., Handley, S. J., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*, *77*(3), 197–213. https://doi.org/10.1016/S0010-0277(00)00098-6

Gelman, S. A., & Kremer, K. E. (1991). *Understanding Natural Cause: Children’s Explanations of How Objects and Their Properties Originate p m^*. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8624.1991.tb01540.x

Gelman, S. A., Raman, L., & Gentner, D. (2009). Effects of Language and Similarity on Comparison Processing. *Language Learning and Development*, *5*(3), 147–171. https://doi.org/10.1080/15475440902824079

Göksun, T., George, N. R., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Forces and Motion: How Young Children Understand Causal Events. *Child Development*, *84*(4), 1285–1295. https://doi.org/10.1111/cdev.12035

Heit, E. (2000). *Features of Similarity and Category-Based Induction*. Retrieved from https://pdfs.semanticscholar.org/8d43/8f1d7b37c4aa05cde34b098376ee4af8c1bf.pdf

Hoyos, C., Horton, W. S., & Gentner, D. (2015). *Analogical comparison aids false belief understanding in preschoolers*. Retrieved from http://groups.psych.northwestern.edu/gentner/papers/HoyosHorton&Gentner_2015.pdf

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251. https://doi.org/10.1037/h0034747

Kalish, C. (1998). Young children’s predictions of illness: Failure to recognize probabilistic causation. *Developmental Psychology*, *34*(5), 1046–1058. https://doi.org/10.1037/0012-1649.34.5.1046

Kjønniksen, L., Torsheim, T., & Wold, B. (2008). Tracking of leisure-time physical activity during adolescence and young adulthood: a 10-year longitudinal study. *The International Journal of Behavioral Nutrition and Physical Activity*, *5*, 69. https://doi.org/10.1186/1479-5868-5-69

Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison promotes learning and transfer of relational categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1303–1310. https://doi.org/10.1037/a0031847

Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by Analogical Bootstrapping. *Journal of the Learning Sciences*, *10*(4), 417–446. https://doi.org/10.1207/S15327809JLS1004new_2

Kushnir, T., & Gopnik, A. (2005). *Young Children Infer Causal Strength From Probabilities and Interventions*. Retrieved from https://journals.sagepub.com/doi/pdf/10.1111/j.1467-9280.2005.01595.x?casa_token=V56i78C9zN8AAAAA:6LEJEQtYh5R0_pldgSmbF9nDFvUdm__SFVVAPlBq5emijI9o3JUhsjAVzNtGAZ-FrSkPQLoy1jCZ4A

Kushnir, T., Xu, F., & Wellman, H. M. (2007). Young children use statistical sampling to infer the preferences of others. https://doi.org/10.1177/0956797610376652

Lagattuta, K. H., & Sayfan, L. (2011). Developmental changes in children's understanding of future likelihood and uncertainty. *Cognitive Development*, *26*(4), 315–330. https://doi.org/10.1016/j.cogdev.2011.09.004

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288. https://doi.org/10.1016/S0010-0277(87)80006-9

Mason, L. (2004). Fostering Understanding by Structural Alignment as a Route to Analogical Learning. *Instructional Science*, *32*(4), 293–318. https://doi.org/10.1023/B:TRUC.0000026512.88700.32

Noyes, A., & Christie, S. (2016). Children Prefer Diverse Samples for Inductive Reasoning in the Social Domain. *Child Development*, *87*(4), 1090–1098. https://doi.org/10.1111/cdev.12522

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-Based Induction. *Psychological Review*, *97*(2), 185–200. Retrieved from http://matt.colorado.edu/teaching/categories/oswls90.pdf

Paavola, M., Vartiainen, E., & Haukkala, A. (2004). Smoking, alcohol use, and physical activity: A 13-year longitudinal study ranging from adolescence into adulthood. *Journal of Adolescent Health*, *35*(3), 238–244. https://doi.org/10.1016/j.jadohealth.2003.12.004

Rubovits, D. S., & Wolynn, T. H. (1999). Children's Illness Cognition: What Mothers Think. *Clinical Pediatrics*, *38*(2), 99–105. https://doi.org/10.1177/000992289903800206

Solomon, G. E., & Cassimatis, N. L. (1999). On facts and conceptual systems: Young children's integration of their understandings of germs and contagion. *Developmental Psychology*, *35*(1), 113–126. https://doi.org/10.1037/0012-1649.35.1.113

Solomon, G. E., Johnson, S. C., Zaitchik, D., & Carey, S. (1996). Like father, like son: young children's understanding of how and why offspring resemble their parents. *Child Development*, *67*(1), 151–171. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8605825

Teglas, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference. *Science*, *332*(6033), 1054–1059. https://doi.org/10.1126/science.1196404

Tversky, A., & Kahneman, D. (1974). *Judgment under Uncertainty: Heuristics and Biases*. *New Series* (Vol. 185). Retrieved from http://www.its.caltech.edu/~camerer/Ec101/JudgementUncertainty.pdf

Wedell, D. H. (2011). Probabilistic Reasoning in Prediction and Diagnosis: Effects of Problem Type, Response Mode, and Individual Differences. *Journal of Behavioral Decision Making*, *24*, 157–179. https://doi.org/10.1002/bdm.686

White, P. A. (1988). Causal processing: Origins and development. *Psychological Bulletin*, *104*(1), 36–52. https://doi.org/10.1037/0033-2909.104.1.36

Appendix

Study 2 Stimuli

Learning:

1. Amy ate a lot of candy. Then, the candy made Amy get a really bad stomachache and she had to stay home from school. Then, Mark came over to play with Amy. Two days later, Mark didn't get sick. (event-stomachache)
2. Amy ate food with germs on it. Then, the germs made Amy get a really bad stomachache and she had to stay home from school. Then, Mark came over to play with Amy. Two days later, Mark got sick. (germ-stomachache)
3. George breathed in some smoke. Then, the smoke made George get a really bad cough and he had to stay home from school. Then, Rachel came over to play with George. Two days later, Rachel didn't get sick. (event-cough)
4. George breathed in some germs. Then, the germs made George get a really bad cough and he had to stay home from school. Then, Rachel came over to play with George. Two days later, Rachel got sick. (germ-cough)

Test:

1. Beth sat out in some flowers for a long time. Then, the flowers made Beth sneeze a lot and she had to stay home from school. Then, Joey came over to play with Beth. Three days later, do you think Joey got sick, or do you think Joey didn't get sick? (event-sneeze)
2. Beth played with a toy with some germs on it. Then, the germs made Beth sneeze a lot and she had to stay home from school. Then, Joey came over to play with Beth. Three days later, do you think Joey got sick, or do you think Joey didn't get sick? (germ-sneeze)
3. John sat in the sun for a long time. Then, the sun made John get a really bad rash and he had to stay home from school. Then, Ellen came over to play with John. Three days later, do you think Ellen got sick, or do you think Ellen didn't get sick? (event-rash)
4. John sat in poison ivy for a long time. Then, the sun made John get a really bad rash and he had to stay home from school. Then, Ellen came over to play with John. Three days later, do you think Ellen got sick, or do you think Ellen didn't get sick? (germ-rash)

Practice and Catch Trials:

1. Here's a cup. And here's a pitcher, and the pitcher has water in it, see? Then, we pour water from the pitcher into the cup. After we pour the water from the pitcher into the cup, does the cup look like this [full cup] or this [empty cup]?
2. Here's a glass cup, and it's on a table. Then, the glass cup starts to get pushed off the table, see? Then, the glass cup falls off the table. After the glass cup falls off the table, does the picture look like this [cup upright on the ground] or this [shattered cup on the ground]?