Children use probability to infer other people's emotions

by

Tiffany Doan

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Hyowon Gweon Associate Professor Department of Psychology Stanford University
Supervisor	Stephanie Denison Associate Professor Department of Psychology University of Waterloo
Internal Member	Ori Friedman Professor Department of Psychology University of Waterloo
Internal Member	Katherine White Associate Professor Department of Psychology University of Waterloo
Internal-external Member	Mathieu Doucet Associate Professor Department of Philosophy University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

I am either the majority contributor or sole contributor to all of the work presented in this thesis. This includes Chapters 2, 3, and 4 of my thesis which are publications co-authored with my supervisor Dr. Stephanie Denison, and collaborator Dr. Ori Friedman. Citations and information regarding these publications can be found in the relevant chapters.

Abstract

People's emotional reactions often depend on probability. However, it is unknown whether children consider probability when inferring other people's emotions. Across three papers, this dissertation shows that children (N = 1465) and adults (N = 481) use probabilistic information when inferring emotions and that this ability develops with age. Chapter Two examined whether children use probability when inferring other people's surprise (four experiments). When inferring who would be surprised about getting a red gumball, 7-year-olds inferred that the person who had a lower chance of receiving a red gumball would be surprised, but younger children did not. Six-year-olds' surprise inferences improved when they were prompted to consider probability, but not when prompted to consider others' prior beliefs. Together, the findings from this chapter show development in children's ability to use probability to infer surprise. Chapter Three examined whether children and adults use probability when inferring other people's happiness (five experiments). When judging the quality of an outcome of two yummy and two yucky gumballs, by age 4, children judged that this outcome is better if it came from a gumball machine that contained mostly yucky gumballs than if it came from a machine that contained mostly yummy gumballs. However, it is not until age 5 that they also recognize that people would be happier with this outcome if it came from the former machine rather than the latter machine. Together, the findings from this chapter show development in children's ability to use probability to infer happiness and a developmental lag between children's assessments of quality and happiness. Chapter Four examined whether children and adults consider close counterfactual alternatives (that are manipulated through probability) when inferring other people's happiness and whether they recognize that an event had a close counterfactual alternative (four experiments). When inferring a person's happiness

about choosing a regular balloon (on a blind choice) from ten closed boxes, 6-year-olds inferred that the person would be happier if they later learned that most of the remaining boxes also held regular balloons than if they mostly held special balloons. However, when asked explicitly about the counterfactual alternative, they did not recognize that a special balloon could have easily been obtained when most of the boxes contained them. Younger children did not make either inference. Children's happiness inferences improved when extra cues were provided, such that 5- and 6-year-olds inferred that a person would be happier when they later learned that they had been physically close to many regular balloons compared to when they were physically close to many special balloons. Six-year-olds also acknowledged that the special balloon was a close counterfactual alternative when the special balloons were physically close and more numerous. Together, the findings from this chapter show that children either infer happiness without considering close counterfactual alternatives, or that counterfactuals influence children's happiness inferences before they can explicitly acknowledge their closeness. As a whole, this dissertation provides evidence that children use probabilistic information to infer other people's emotions.

Acknowledgements

There are so many people that I would like to thank for supporting me throughout this journey. First, I would like to extend my sincere gratitude to my supervisor, Stephanie Denison. Thank you for giving me the freedom to explore my interests, for supporting and mentoring me for these past six years, for helping me become a better thinker, writer, and researcher, and for always believing in me and encouraging me to step out of my comfort zone. I honestly cannot explain how grateful I am to be your student.

To the friends that I met on day one and the ones I met along the way (you know who you are), thank you for being on this journey with me. These past six years would have been much more difficult without our office drop-by chats, SLC runs, random walks, late-night basketball games, softball games, skating classes, game nights, and random hangouts. I did not expect to meet so many amazing people, but I am so thankful to have you all in my life.

To my parents and sister, your unconditional love and support is what got me to the finish line. Mom and dad, I owe you everything. I am who I am today because of you. Jenn, thank you for always believing in me, even when I didn’t believe in myself. And Will, words cannot describe how grateful I am to have you in my life. Thank you for always being there for me. You are my rock.

Dedication

This dissertation is dedicated to my mom, dad, and sister. We did it.

Table of Contents

List of Figures

List of Tables

Chapter One: General Introduction

Probability influences people's emotions. For example, say your friend won $100 from a contest. You would probably infer that she would be extremely surprised about her win if 500 other people had entered the contest. However, you would infer that she would not be very surprised if only one other person had entered the contest. Similarly, you might infer that your friend would be much happier about winning in the former scenario rather than the latter, as her chances of winning were initially low and so she might be *pleasantly* surprised. Thus, the same outcome can elicit different emotions or degrees of an emotion depending on the chances of that outcome occurring.

The ability to infer others' emotions is crucial for successful social interactions. Knowing how and why someone feels a particular way allows us to explain their past behaviors, predict their future actions, and even helps us regulate our own behaviors. Similarly, failing to correctly infer others' emotions can impair our ability to empathize with or to appropriately comfort them. These abilities are especially important in childhood because difficulties in emotion attribution could significantly disrupt children's social interactions and interfere with their ability to form friendships. As such, it is important to investigate the factors that children use when inferring people's emotions. This dissertation investigates one such factor that has been mostly neglected in the developmental literature – namely, whether children consider probabilistic information when inferring other people's emotions.

Before examining whether children integrate their understanding of probability with their understanding of emotions, it is important to know what children understand about each of these concepts individually, and whether adults consider probability in their emotion inferences. As such, in this chapter, I first discuss infants' and young children's ability to engage in probabilistic reasoning and how they use probability in their social interactions. Second, I discuss children's development in emotion understanding and the current theories of how children infer emotions. Next, I review the literature on adults' use of probability when inferring surprise and happiness. Finally, I provide an overview of the empirical studies conducted in the following chapters.

What do we know about children's probabilistic reasoning?

Numbers and proportions are all around us and we use probability information in our daily lives, whether we explicitly recognize it or not. Whether we bring an umbrella to work depends on the chances that it will rain. A baseball coach's decision for his lineup depends on his players' batting averages. A teacher's decision to accept a student's excuse might depend on how reliable they have been in the past. As such, information about statistics, proportions, and probabilities help us navigate the world around us. In fact, even infants are sensitive to this information as early as their first year of life (e.g., Denison et al., 2013; Téglás et al., 2007; Wellman et al., 2016; Xu & Denison, 2009; Xu & Garcia, 2008).

In violation-of-expectation (VOE) looking-time paradigms, infants look longer at events that they find "surprising" or unexpected. This logic has been applied to study probabilistic reasoning in infancy. By 6 months of age, infants will look longer at a sample of 4 yellow and 1 pink balls than a sample of 4 pink and 1 yellow balls if the sample came from a larger population that contained a ratio of 4 pink to 1 yellow balls (Denison et al., 2013). This suggests that they expect a sample to reflect its population and that their attention is drawn to samples that violate the statistical properties of the population from which it is drawn. Infants can also make the reverse inference: when they are presented with five randomly sampled items from a population, they expect that the larger population will have similar proportions to the sample (Xu & Garcia, 2008). By 12 months of age, infants will look longer when a single minority object exits a container than when one of three majority objects exits, demonstrating that they have expectations about future single-sampling events based on the initial likelihood of those events (Téglás et al., 2007). These findings suggest that infants have an understanding of the statistical relationship between a sample and the population that it came from.

However, infants do not automatically expect that a sample should statistically match its population, and in fact, they integrate a number of factors when determining the relationship between a sample and a population. When 11-month-old infants are shown that an experimenter has a preference or goal for white over red balls and that she is intentionally and not randomly sampling, they expect that she will sample white (and not red) balls from a population, even if the population has mostly red balls. However, if the experimenter is blindfolded and cannot see the population that she is sampling from, then infants expect her to sample mostly red balls, matching the population's distribution (Xu & Denison, 2009). This suggests that even infants are sensitive to a number of factors (e.g., preference, visual access) that influence our statistical inferences and can use that information to make inferences.

Infants and young children can also integrate their intuitions about probability with their understanding of the *social* world. When learning from others, infants consider both the sample and the sampling process when making inductive generalizations (Gweon et al., 2010). In one study, 15-month-old infants saw an experimenter squeak three blue balls after sampling them from either a population of mostly blue balls and few yellow balls, or a population of mostly yellow balls and few blue ones. When given the chance to play with a yellow ball, infants squeezed the yellow ball *less* when the blue balls were sampled from the latter population (Gweon et al., 2010). This suggests that they recognized that the experimenter purposely violated random sampling in the population of mostly yellow balls in order to teach them about which toys squeaked. Thus, infants constrained their generalization of the squeaking feature to the blue balls when there were few of them in the population but not when they were plentiful. By 20 months of age, probability influences toddlers' and preschoolers' explicit social inferences, such as when inferring another person's preferences (Diesendruck et al., 2015; Kushnir et al., 2010; Ma & Xu, 2011). For example, when a person pulls a few duck toys from a box containing duck and frog toys, young children infer that the person has a preference for ducks if the box contains mostly frog toys, but not if the box contains mostly duck toys (Kushnir et al., 2010). This inference relies on probability. Pulling only duck toys from a mostly frog population violates random sampling and suggests that the person intentionally chose only duck toys because they have a preference or goal for ducks over frogs.

As a whole, the research discussed in this section shows that young children can use sampling behaviour and probabilistic information to make inferences about other people and their environment. Given that even infants are sensitive to probability information and that young children use it in at least one kind of social inference, it is plausible that children might also use probability when inferring emotions.

What do we know about children's emotion understanding?

Children's emotion understanding is an important aspect of their social cognitive development and has been linked with a variety of positive outcomes. For example, children with better emotion understanding are more successful in their social interactions (e.g., Cassidy et al., 1992; McDowell et al., 2000), have fewer anxiety and depressive symptoms (e.g., Rieffe & de Rooij, 2012; Rieffe et al., 2007, 2008), and perform better academically (e.g., Jones et al., 2011;

Lecce et al., 2011). As such, much research has been devoted to examining the development of children's emotion understanding.

Children are exposed to and experience emotions from birth and their emotion understanding develops rapidly throughout childhood. At 3 months of age, infants can recognize and differentiate smiling from frowning faces (Barrera & Maurer, 1981), and at 5 months, they can distinguish between happy and sad vocalizations (Walker-Andrews & Lennon, 1991). Their ability to differentiate between emotions continues to develop until the toddler years when they are able to recognize and name basic emotions such as happiness, sadness, fear, and anger (e.g., Bullock & Russell, 1985; Denham, 1986; Hughes & Dunn, 1998). Around age 3 to 4, children begin to understand how external causes can affect other people's emotions (e.g., Barden et al., 1980; Borke, 1971; Harris et al., 1987), and by age 6, they can appreciate that people's emotions depend on their desires and beliefs (e.g., Hadwin & Perner, 1991; Harris et al., 1989). By age 8, children develop more complex understandings of emotions such as the understanding that people can have multiple or mixed emotions toward one situation (e.g., Donaldson & Westerman, 1986; Kestenbaum & Gelman, 1995). While the development of emotion understanding is important as a whole, this dissertation focuses specifically on children's *inferences* of other people's emotions.

There are currently two leading accounts of how children infer emotions. One account proposes that children infer emotions by relying on memorized scripts (e.g., Barden et al., 1980; Gove & Keating, 1979; Harris et al., 1987; Hughes et al., 1981; Widen & Russell, 2010, 2011; also see Fehr & Russell, 1984). Scripts are knowledge structures consisting of sequences of specific, concrete events (e.g., Abelson, 1981; Schank & Abelson, 1975) that allow people to follow social conventions and to reason about common events. For example, our ability to order food at a restaurant depends on our having learned a script outlining the sequence of events that unfolds when ordering at a restaurant (waiter brings menu, customer decides what to order, customer tells waiter, waiter tells cook; Abelson, 1981). Applied to children's understanding of emotions, the script theory suggests that with development, children come to learn the antecedents of particular emotions. Once these scripts are learned, children can use them to infer other people's emotions. For example, children may learn the script that people are happy when they receive presents. When they are going to a friend's birthday party, they know that their friend will receive gifts, so they can predict that their friend will be happy (e.g., Widen & Russell, 2011).

Another account suggests that children infer emotions by considering people's mental states, like their desires, beliefs, and goals (e.g., Harris, 2008; Lagattuta, 2005, 2008; Rieffe et al., 2005; Skerry & Spelke, 2014; Wellman & Banerjee, 1991; Wellman & Bartsch, 1988). By the age of 2, children consider people's desires and whether or not those desires are fulfilled to infer whether they will be happy or sad. For example, children predict that a boy who is searching for his dog will be happy if he finds it, but sad if he does not (Wellman & Woolley, 1990). With age, children also begin to consider people's beliefs when inferring their emotions, though the age at which they do so varies across emotions. For example, at around age 5, children start to consider people's beliefs to infer whether they will be happy or sad (e.g., Hadwin & Perner, 1991; Harris et al., 1989). They predict that a boy who likes candy will be happy if he *believes* a box contains some, even if the box has no candy and the actual contents are undesirable to him (Hadwin & Perner, 1991). As they get older, children also consider people's beliefs when inferring their surprise (age 7, MacLaren & Olson, 1993; Ruffman & Keenan, 1996) and fear (age 6 or 7, e.g., Bradmetz & Schneider, 1999; Ronfard & Harris, 2014).

It is easy to think about how the script and theory of mind accounts work for simple scenarios like "getting presents makes people happy". But how do we infer the subtler and more sophisticated parts of people's emotional lives? How do we infer the degree of people's emotions? Or deal with instances where people have the same desires or goals and experience identical events but do not feel the same emotion? What do we do when multiple cues to emotions are at play? It is harder to imagine how these theories can account for these more complex cases. This raises the question of what other factors matter for emotion cognition.

One element of our more general reasoning abilities that may play a role is probabilistic or statistical reasoning. Information about probability influences children's learning and inferences in many areas of non-social and social reasoning. For example, young infants use transitional probability to segment continuous streams of speech syllables into words, non-linguistic auditory sequences into 'tone words', and strings of visual shapes into pairs (Aslin et al., 1998; Kirkham et al., 2002; Saffran et al., 1996; Saffran et al., 1999). Preschoolers make statistical inferences and consider the sampling processes that generate the labels of objects when generalizing novel words (Xu & Tenenbaum, 2007a, 2007b). They also use and combine information about prior probability and the conditional probability of events to make judgments about causation (Gopnik & Schulz, 2004; Gopnik et al., 2001; Sobel et al., 2004). While less work has examined how children's probabilistic reasoning matters in the social world, previous work has shown that when learning from others, infants consider both the sample and the sampling process when making inductive generalizations (Gweon et al., 2010), and young children consider probabilistic information when making explicit inferences about people's preferences (e.g., Kushnir et al., 2010; Ma & Xu, 2011). As such, statistical and probabilistic reasoning may also play a role in children's emotion cognition.

In fact, reasoning about statistics and probability may be well suited for understanding the more sophisticated, unpredictable, and complex aspects of people's emotional lives. Information about probability can help us infer the degree of emotion someone feels and can help explain why two people can have similar mental states and experience identical outcomes in a situation, and still feel differently. For example, if somebody receives a positive outcome, they might feel happier about it if there was a low prior probability of the outcome occurring compared to a high prior probability. Similarly, two people who have the same desires and experience the exact same outcomes may feel differently depending on their initial chances of receiving that outcome. Further, there are instances where we do not have access to other people's mental states or cases where people struggle with theory of mind reasoning (e.g., individuals with autism), so being able to consider other factors like probability to infer emotions can be beneficial. Thus, the current dissertation investigates the development of children's ability to use probability information when inferring other people's emotions, specifically, people's surprise and happiness.

Examining the development of children's ability to integrate their understandings of these two seemingly independent concepts will broaden our knowledge of human cognition and will provide us with a more extensive understanding of children's emotion inferences. Using probability to infer emotions cannot be fully accounted for by memorizing scripts or mental state reasoning, so if children use probability to infer emotions, theories about children's inferences of emotions should be revised to account for factors like probability. Specifically, theories of emotion should be able to account for the numerous, complex, and sometimes divergent cues that people encounter in the social world, and how they integrate these multiple cues to accurately infer others' emotions. A recently proposed framework for studying how people reason about emotions may do just this (Ong et al., 2015, 2019). Using a computational, lay theory approach, this framework examines how people reason and combine information from multiple sources to infer emotions and how people use domain-general processes (similar to other forms of cognition and social cognition) to reason about emotion. This intuitive theory framework of emotion may be fruitful for integrating probabilistic reasoning into children's conceptions of emotion. I will return to this point in the General Discussion.

Adults use probability to infer emotions

While classic work with adults has examined how probability information influences emotion inferences (e.g., Bell, 1985; Kahneman & Tversky, 1982; Kahneman & Varey, 1990; Mellers et al., 1997; Meyer et al., 1997), there is still much to be learned. In this section, I give a brief summary of such work with adults and discuss how the findings from the current dissertation could add to this literature.

The intuitive view of surprise is that it arises when unexpected events occur, and that expectations can be conceptualized in terms of probabilities (Meyer et al., 1997; Teigen & Keren, 2003). An unlikely event is surprising, but a likely event is not. Adults infer surprise in this way. For example, when asked to make probability and surprise judgments about an occurrence of rainfall, adults are more surprised when the probability of rain was low compared to high (Maguire et al., 2011). They also make this inference for others. When judging how surprising it would be for someone to catch their connecting flight after their first flight was delayed, adults say that it would be more surprising if the person initially only had a 40% chance of making it compared to if they initially had a 60% chance (Teigen & Keren, 2002). This suggests that adults understand the link between probability and surprise.

Although less intuitive than the connection between probability and surprise, people's happiness can also be influenced by probability. In gambling games, adults are happier about a win if their initial chances of winning were low compared to high and are more disappointed with a loss if their initial chances of winning were high as compared to low (Mellers et al., 1997; van Dijk & van der Pligt, 1997). This difference in happiness towards the same outcome is thought to depend on counterfactual reasoning, in which people compare actual outcomes with possible alternative outcomes (e.g., Bell, 1985; Loomes & Sugden, 1986; Shepperd & McNulty, 2002). That is, a positive outcome feels better when a worse alternative outcome was more likely, and a negative outcome feels worse when a better alternative outcome was more likely.

In the current dissertation, I first examine whether children use probability to infer surprise (Chapter Two). This will allow us to examine the development of conceptions of surprise, by assessing whether children's and adults' conceptions of surprise are similar. Next, I examine whether children and adults use probability to infer happiness (Chapter Three) and whether these inferences are dependent on counterfactual reasoning (Chapter Four). It may seem redundant to investigate these questions with adults because there is existing work on this topic. However, the design of the current studies show the odds of events using clear, easily enumerable, visual paradigms, which has not been done in the past. Most research with adults presented the scenarios in relatively complex vignettes (e.g., Shepperd & McNulty, 2002; van Dijk & van der Pligt, 1997), which do not lend themselves as well to clearly quantified outcomes. Further, the findings from my studies with children and adults can be compared directly as the same visual scenarios are used to assess their understanding.

Overview of Dissertation

The following chapters describe several empirical studies examining children's (and adults') consideration of probability when inferring others' emotions. Chapter Two investigates the development of 4- to 7-year-old children's use of probability in inferring surprise and two manipulations intended to make the connection between probability and surprise more apparent to younger children. In the first three experiments, children saw stories where two characters received a red gumball from different gumball machines. One machine contained mostly red gumballs and the other machine contained only a minority of red gumballs. Children were asked which character was more surprised to get a red gumball. Experiment 1 examines the development of children's ability to use probability when inferring surprise. Experiment 2 examines whether children's surprise inferences would improve when the events are deterministic and one event is impossible. Experiment 3 looks at whether their surprise inferences improve when prompted to consider probability and people's beliefs. Experiment 4 replicates the findings of Experiment 3, using a slightly different design in which children attributed emotions to a single character.

Chapter Three investigates the development of 4- to 6-year-old children's use of probability in inferring happiness and the quality of outcomes. In five experiments, children and adults saw stories where a girl received two red and two black gumballs from a gumball machine. In one condition, the gumball machine contained mostly red yummy gumballs and a few black yucky ones. In the other condition, the machine contained mostly black yucky gumballs and a few red yummy ones. Children and adults were asked to rate how the girl felt or how good the outcome was. Experiments 1 and 2 examine the development of children's ability to consider probability when inferring happiness. Experiment 3 looks at the development of children's ability to use probability when assessing the quality of an outcome. Experiment 4 examines the developmental lag between children's ability to use probability to assess quality and happiness. Experiment 5 examines adults' use of probability when inferring outcome quality and happiness.

Chapter Four further investigates the development of 4- to 6-year-old children's use of probability in inferring happiness, while also asking whether they do so by considering close counterfactual alternatives. In four experiments, children and adults saw stories about a girl who won a mundane prize. In one condition, the girl later discovered that her odds of winning a better prize had been high. In the other condition, the girl later discovered that her odds of winning a better prize had been low. Children and adults were asked to rate how the girl felt and were asked a question that assessed whether they recognized the closeness of the counterfactual alternative of the girl winning the better prize. Experiments 1 and 2 examines whether children consider the initial odds of winning a better prize when inferring happiness and whether they recognize the closeness of the better prize. Experiment 3 looks at whether children's judgments would improve when additional cues to the closeness of the alternative were provided. Experiment 4 examines adults' consideration of probability and close counterfactual alternatives when inferring happiness.

Chapter Two: Children use probability to infer other people's surprise (Paper One)

A version of this paper is published:

Doan, T., Friedman, O., & Denison, S. (2018). Beyond belief: The probability-based notion of surprise in children. *Emotion, 18*(8), 1163-1173. doi: 10.1037/emo0000394

Probability and surprise go hand in hand – improbable events are surprising, but probable ones are not. As the example in Chapter One illustrates, you would infer that your best friend would be surprised to win $100 if 500 other people had entered the contest, but not if only one other person had entered. This connection between probability and surprise is present in adults' conceptions of surprise, as they attribute surprise to agents who observe improbable outcomes (e.g., Maguire et al., 2011; Teigen & Keren, 2003).

However, it is unknown whether children consider probability when inferring surprise. Previous research has not examined this, as most research on children's understanding of surprise relates it to their understanding of others' beliefs. When children explain why a character was surprised by an outcome, 4-year-olds refer to the character's beliefs, though 3-year-olds refer to the character's desires (Wellman & Banerjee, 1991; Wellman & Bartsch, 1988). For example, 4-year-olds might say a boy is surprised that his grandmother's house is purple because he thought the house would be white, whereas 3-year-olds might say he is surprised because he likes purple (Wellman & Banerjee, 1991). Further, 4- and 5-year-olds appropriately attribute surprise to characters whose beliefs are not met, but refrain from attributing surprise to characters whose desires are not met (Hadwin & Perner, 1991; Wellman & Bartsch, 1988). Finally, children who correctly infer another's beliefs are also better at predicting their surprise. For example, when 3- to 8-year-olds judge which of two boxes will surprise a puppet, only children aged 5 and up, who correctly infer the puppet's beliefs about the contents of the boxes, choose the correct box. This finding suggests that in order to understand surprise, children must first understand beliefs (MacLaren & Olson, 1993; also see Ruffman & Keenan, 1996, and Scott, 2017 for conflicting evidence about whether inferring surprise from false belief emerges later or earlier in development).

Children may also use probability to infer surprise. This is plausible, as adults infer surprise in this way (e.g., Maguire et al., 2011; Teigen & Keren, 2003). For example, when asked to make probability and surprise judgments about an occurrence of rainfall, adults rated their surprise as greater when the probability of rain was lower (Maguire et al., 2011). It is important to note that such probability-based inferences may not require attributing beliefs. Adult participants could have attributed surprise by only considering the probability of rain occurring, and without attributing any beliefs to themselves regarding whether it would rain. Thus, belief understanding might not be the *only* requisite for inferring surprise.

If children also use probability to infer surprise, this will advance knowledge of how children understand surprise, and how they infer emotions more broadly. Existing accounts suggest that children understand and infer emotions by memorizing scripts or by considering others' mental states (e.g., Barden et al., 1980; Harris et al., 1989, 2016; Wellman & Woolley, 1990; Widen & Russell, 2010, 2011). Using probability to infer surprise does not fit under either theory. Scripts are composed of specific concrete events (e.g., Abelson, 1981), and do not make reference to underlying abstract concepts, like probability (for further discussion see Gopnik & Meltzoff, 1997, p. 62). Likewise, notions of probability are not included in children's notions of others' mental states. Hence, if children also use probability to infer surprise, this will reveal another method of understanding emotions, as it will show that their understanding of surprise depends on their probabilistic reasoning.

Investigating whether children use probability when attributing surprise will also advance our understanding of probabilistic reasoning in children. Children consider probability information from early in development. Infants use probability to guide their own expectations and slightly older children use probability in social inferences (e.g., Denison et al., 2013; Kushnir et al., 2010; Ma & Xu, 2011; Téglás et al., 2007). If children also use probability to infer emotions, this will expand our knowledge of its influence in children's social cognition. Further, investigating how this ability develops could be informative about how children relate and integrate concepts from different domains.

In four experiments, we investigate children's ability to use probability to explicitly attribute surprise to another person. In the first three experiments, children saw stories in which two characters received a red gumball from different gumball machines. One machine contained mostly red gumballs and the other machine contained only a minority of red gumballs. Children were asked which character was more surprised with the outcome. In the fourth experiment, we used a slightly different method, in which children attributed emotions to a single character. These four experiments investigate the development of children's use of probability in inferring surprise and two manipulations intended to make the connection between probability and surprise more apparent to younger children.

Experiment 1

Experiment 1 investigated the development of children's ability to use probability to infer surprise. Children were told stories about two characters at two different gumball machines and were asked to choose the character who was more surprised after seeing both characters receive a red gumball. To succeed, children had to appreciate that although getting a red gumball was probable for one character, it was improbable for the other character, making the outcome surprising.

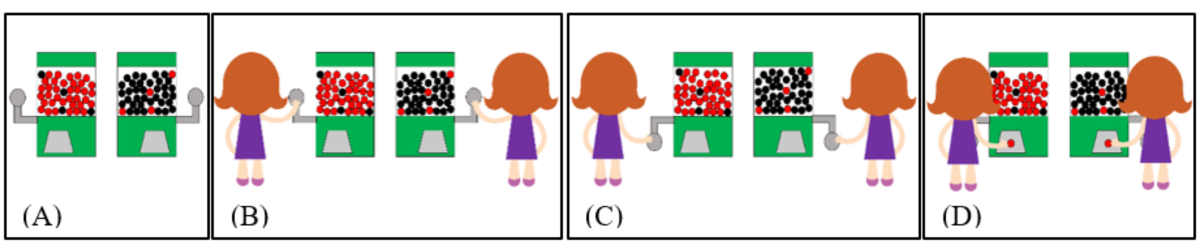
Method

Participants. One hundred and twenty children participated: 30 4-year-olds (*M* = 4;7 [years; months]; range = 4;3 – 4;11; 15 girls), 30 5-year-olds (*M* = 5;5; range = 5;0 – 5;11; 14 girls), 30 6-year-olds (*M* = 6;6; range = 6;0 – 6;11; 13 girls), and 30 7-year-olds (*M* = 7;4; range = 7;0 – 7;11; 13 girls). In all experiments, children were individually tested at schools and daycares in the Waterloo Region. Demographic information was not formally collected, but the region is predominantly middle-class, and approximately 79% of residents in this region are Caucasian, with Chinese and South Asians as the most visible minority. Different children were tested in each experiment. This research, submitted under the name, "Social Understanding in Children" (ORE#20042), received ethics clearance through the University of Waterloo's Research Ethics Committee.

Materials and procedure. All materials in the current experiment and the following experiments were shown on a laptop computer. Children were told two stories (2 trials). Each story was about two gumball machines. One machine contained many red gumballs and just a few black ones (36 red, 4 black), and the other machine contained the reverse distribution (36 black, 4 red). We used this distribution (9:1) to ensure that children would readily notice that one colour was more plentiful than the other (e.g., Denison et al., 2013; Girotto et al., 2016). Children were told that the red gumballs are yummy and the black gumballs are yucky. Two identical-looking characters, always viewed from behind, then appeared, with each character at one machine. The characters were depicted this way to prevent children's responses from being swayed by extraneous factors, such as differences between the characters, or the expressions on their faces. Children were told that both characters wanted a red gumball and were asked a comprehension check question to confirm that they understood (i.e., "What colour gumball do the girls want?"). The characters pulled the handles of their machines, and each ended up getting a red gumball. Children were then asked which character was more surprised. See Figure 1 for a sample story and script.

In the first story, the characters were girls, the gumball machines were coloured green, and they appeared side-by-side. In the second story, the characters were boys, the machines were coloured orange, and one appeared above the other. We varied these details to prevent children from repeating or alternating responses across the stories. The location of the gumball machines was counterbalanced across participants, such that for half the children the mostly-red machine was on the right in the first story and on the top in the second story, and in the opposite locations for the other children.

If children responded incorrectly to the comprehension check question about which type of gumball the characters wanted, the experimenter repeated the information about the tastes of the red and black gumballs. When the comprehension check question was asked again, all children answered correctly.



(A) Here are two gumball machines. They have red gumballs and they have black gumballs. The red gumballs are very, very yummy, and the black gumballs are yucky. (B) And look, here are two girls. They both want a red gumball. *What color gumball do the girls want?* (C) To get a gumball, the girls pull down the handles, and the machines shake up all the gumballs. (D) Look! They both got a red gumball. So now I have a question for you. *Which girl is more surprised that she got a red gumball?*

Figure 1. Sample slides and script from Experiment 1 (Trial 1).

Results and Discussion

Six children initially responded incorrectly to the comprehension check question about which type of gumball the characters wanted (four 4-year-olds; one 5-year-old; one 7-year-old). When these children's data were excluded, our main pattern of results remain the same. Thus, we report our analyses with our full sample.

Of primary interest was whether children are able to use probability to infer the characters' surprise, and whether this ability develops with age; Figure 2 shows children's mean number of correct responses (i.e., choosing the character who was at the machine with fewer red gumballs). A Generalized Estimating Equations (GEE) binary logistic regression with age as a between-subject factor (4, 5, 6, 7) revealed a significant main effect of age, *Wald* $X^2(df = 3, N = 120) = 13.02$, $p = .005$. Pairwise comparisons revealed that 4-year-olds gave significantly fewer correct responses compared to children at all other ages, *p*s ≤ .019. However, responses did not significantly differ between 5-, 6-, and 7-year-olds, *p*s ≥ .214.

We then used Wilcoxon sign tests to examine whether children at each age chose the correct character more or less than half the time (i.e., whether the scores departed from chance score of 1). For these analyses, we summed children's correct responses across both trials for a maximum score of two. Seven-year-olds predominantly chose the correct character ($M = 1.43$, $SD = .817$), $z = -2.60$, $p = .009$, 6-year-olds showed a trend in this direction ($M = 1.27$, $SD = .785$), $z = -1.79$, $p = .074$, 5-year-olds responded at chance ($M = 1.17$, $SD = .874$), $z = -1.04$, $p = .297$, and 4-year-olds predominantly chose the wrong character ($M = 0.67$, $SD = .802$), $z = -2.13$, $p = .033$.

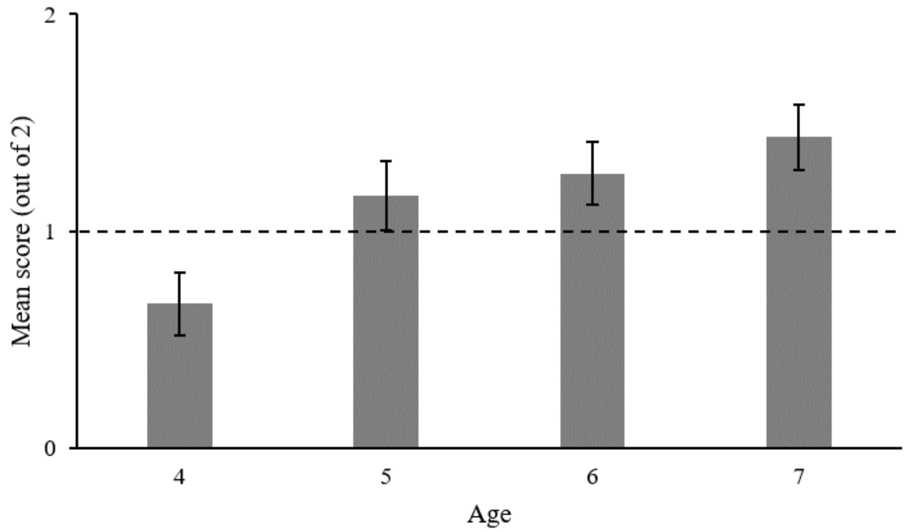


Figure 2. Mean scores for children's surprise judgments in Experiment 1. Error bars show ± 1 standard error of the mean. Dotted line at a score of 1 represents chance performance.

These findings suggest development in children's ability to use probability to infer another's surprise. Although 7-year-olds correctly judged that the character with a lower chance of getting a red gumball was more surprised (and 6-year-olds trended in this direction), 5-year-olds did not make systematic judgments, and 4-year-olds incorrectly judged that the character with a higher chance of getting a red gumball would be more surprised.

One possible interpretation of these results is that 4- and 5-year-olds are insensitive to the probabilities of the distributions in this experiment. However, this is unlikely, as 4-year-olds attributed greater surprise to the character with the more probable distribution, suggesting that even these youngest children were sensitive to the differences between the distributions. Furthermore, it is unlikely that 4-year-olds mistakenly think that the more probable distribution is less likely to yield a red gumball, as previous research suggests that children this age expect the majority item in a distribution to be sampled most often (e.g., Denison et al., 2006, 2013; Girotto et al., 2016).

Another possible explanation for children's poor performance is that they may not yet understand the word "surprise", or may mistakenly think that it means a positive emotion. Some evidence suggests that young children associate surprise with desirable events (e.g., Bartsch & Estes, 1997), attribute positive emotions to the word surprise (Russell, 1990), and know that achieving a goal results in positive emotions (e.g., Skerry & Spelke, 2014; Wellman & Woolley, 1990). However, we tested children aged 4 and older, and previous research shows that children at these ages do understand "surprise", and do not just connect it with positive outcomes (e.g., Hadwin & Perner, 1991; MacLaren & Olson, 1993; Wellman & Banerjee, 1991). For example, 4-year-olds explain surprise by referring to agents' beliefs; a tendency to interpret "surprise" as referring to positive outcomes would instead predict that children should only provide desire-based explanations (Wellman & Banerjee, 1991). Hence, children misunderstanding the meaning of "surprise" is unlikely to explain our findings.

Although children are not limited to interpreting "surprise" as *referring* to positive events, they may nonetheless *associate* it with such events, and with situations that are likely to have positive results. Because of this, telling children about the desirability of the gumballs (i.e., red ones are yummy; black ones are yucky) may have negatively impacted their performance. It may have biased them to choose the character who had access to the machine with the larger proportion of "yummy" gumballs, as this character was more likely to have positive results (i.e., getting even more good gumballs). This could explain why 4-year-olds showed below chance performance. Thus, in the remaining experiments, the desirability information was removed.

Experiment 2

We examined whether making the gumball distributions more extreme would improve children's surprise judgments. The experiment included two conditions: improbable and impossible. The improbable condition used the same distribution of gumballs as Experiment 1, to see if the findings would replicate. In the impossible condition, one gumball machine contained only red gumballs and the other machine contained only black gumballs, and so it should have been impossible for one character to get a red gumball. We hoped this distribution would make the improbability of one character getting a red gumball more salient, and maximally surprising. Because 7-year-olds in the first experiment were able to correctly infer surprise using probabilistic distributions, only 4- to 6-year-olds were tested.

Method

Participants. One hundred and twenty children participated: 40 4-year-olds (*M* age = 4;6 years, range = 4;0 – 4;11; 18 girls), 40 5-year-olds (*M* = 5;5; range = 5;0 – 5;11; 22 girls), and 40 6-year-olds (*M* = 6;3; range = 6;0 – 6;10; 17 girls). Children were recruited and tested at daycare centres and schools.

Materials and procedure. Children were told two stories (2 trials) about gumball machines and were randomly assigned to one of two conditions. In the improbable condition, one machine contained many red gumballs and just a few black ones (36 red, 4 black), and the other machine contained the reverse distribution (36 black, 4 red). In the impossible condition, one gumball machine contained all red gumballs and the other machine had all black gumballs. In both conditions, two identical-looking characters appeared, with each character at one machine. Children were told that both characters wanted a gumball. The characters pulled the handles of their machines, and each ended up getting a red gumball. Children were then asked which character was more surprised. The script was identical across both conditions, and was as follows:

Here are two gumball machines. They have red gumballs and they have black gumballs. And look, here are two girls. They both want a gumball. To get a gumball, the girls pull down the handles, and the machines shake up all the gumballs. Look! They both got a red gumball. So now I have a question for you. *Which girl is more surprised that she got a red gumball?*

Results and Discussion

Of interest was whether children would be sensitive to the varying distribution of the gumball machines, and whether their surprise judgments would be more accurate for the impossible outcomes. Figure 3 shows children's mean number of correct responses (i.e., choosing the character who was at the machine with fewer (or no) red gumballs). A GEE binary logistic regression with age (4, 5, 6) and condition (impossible, improbable) as between-subject factors revealed a significant main effect of age, *Wald* $X^2(df = 2, N = 120) = 7.26$, $p = .027$. There was no effect of condition, *Wald* $X^2(df = 1, N = 120) = 2.24$, $p = .134$, and no condition by age interaction, *Wald* $X^2(df = 2, N = 120) = 1.69$, $p = .431$, though it is possible that a significant interaction would be revealed if we tested a larger sample of children. Pairwise comparisons revealed that 6-year-olds performed significantly better than 4-year-olds, $p =.006$, and 5-year-olds, $p =.038$, but 4-year-olds and 5-year-olds did not differ in their performance, $p = .635$.

We then summed children's correct responses across both trials for a maximum score of two. Wilcoxon sign tests collapsed across conditions revealed that 6-year-olds ($M = 1.33$, $SD = .764$) predominantly chose the correct character, $z = -2.50$, $p =.012$, whereas, 4-year-olds ($M = 0.90$, $SD = .632$), $z = -1.00$, $p =.317$, and 5-year-olds ($M = 0.98$, $SD = .768$), performed at chance, $z = -0.21$, $p = .835$.

These results reveal age-related improvements in children's ability to use probability to infer surprise. However, children's judgments of surprise did not improve when the outcome was impossible. Additionally, removing the desirability of the red gumballs moved 4-year-olds' responses closer to chance, suggesting that 4-year-olds' tendency to choose the more probable distribution in Experiment 1 was due to their inability to inhibit the impulse to choose the machine with more "yummy" gumballs. We next examined whether other manipulations might further improve children's ability to infer surprise.

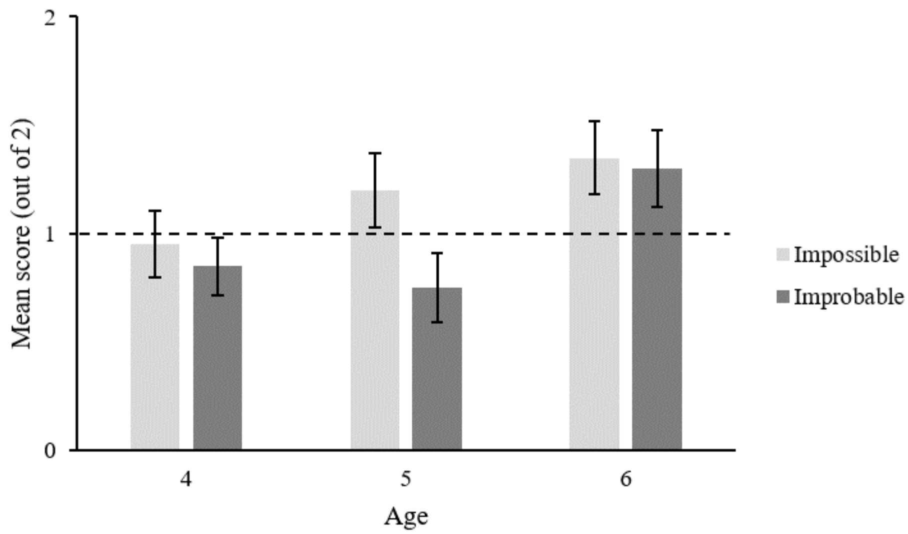


Figure 3. Mean scores for children's surprise judgments in Experiment 2 for the impossible and improbable conditions. Error bars show ± 1 standard error of the mean. Dotted line at a score of 1 represents chance performance.

Experiment 3

Thus far, it appears that young children have difficulty using probability to judge surprise. One explanation for this is that younger children do not see a connection between probability and surprise. Alternatively, they might understand that a connection exists, but might not *spontaneously* consider probability when inferring surprise – they might not spontaneously consider that one girl had a better chance of getting a red gumball than the other. A related explanation for young children's difficulty, though, is that they might not spontaneously consider the characters' beliefs about getting a red gumball – they might not spontaneously consider that one girl believed she would get a red gumball and the other did not. On these views, children might perform better if they were explicitly prompted to consider the characters' chances or beliefs before making a surprise inference.

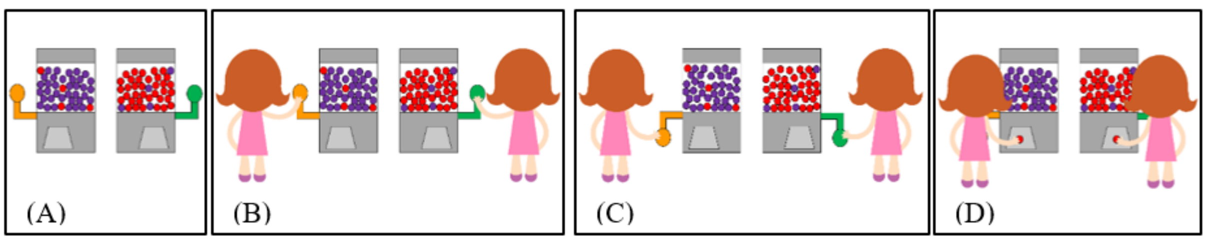
To examine these possibilities, children were asked a prompt question before seeing both characters receive a red gumball. In a belief condition, they were asked which character *thinks* they are going to get a red gumball; in a probability condition, they were asked which character has *a better chance* of getting a red gumball; finally, in a control condition, they were asked which character was at a machine with a particular coloured handle (always corresponding to the mostly red machine, as in the belief and probability conditions).

Method

Participants. One hundred and twenty children participated: 60 5-year-olds (*M* = 5;6; range = 5;0 – 5;11; 27 girls), and 60 6-year-olds (*M* = 6;4; range = 6;0 – 6;11; 31 girls). Children were either tested at schools or in a quiet lab setting.

Materials and procedure. Children were again told two stories (2 trials) about two gumball machines. One machine contained many red gumballs and just a few purple ones (36 red, 4 purple), and the other machine contained the reverse distribution (36 purple, 4 red). Two identical-looking characters then appeared, with each character at one machine. Children were told that both characters wanted a gumball. Children were randomly assigned to one of three prompt conditions. We describe these questions by referring to the scripts with the two girls. The belief prompt asked, "Which girl *thinks* she's going to get a red gumball?", the probability prompt asked, “Which girl has a *better chance* of getting a red gumball?”, and the control prompt asked, “Which girl is standing beside the machine with a green handle?” After the prompt question, the characters pulled the handles of their machine, and both characters received a red gumball. Children were asked which character was more surprised. See Figure 4 for a sample of the story and script for the probability prompt.

A few children initially failed the prompt question (see Results). When this happened, the experimenter said, “Let’s hear the story again”, repeated the story from the beginning, and re-asked the prompt question. The experimenter repeated the prompt a maximum of two times (only four children needed to have the story repeated twice). All children answered the prompt question correctly after the repetitions.



(A) Here are two gumball machines. They have red gumballs and they have purple gumballs. (B) And look, here are two girls. They both want a gumball. *Which girl has a better chance of getting a red gumball?* (C) To get a gumball, the girls pull down the handles, and the machines shake up all the gumballs. (D) Look! They both got a red gumball. So now I have a question for you. *Which girl is more surprised that she got a red gumball?*

Figure 4. Sample slides and script of the probability prompt from Experiment 3 (Trial 1).

Results and Discussion

We first examined children’s *initial* responses to the prompt questions. Wilcoxon sign tests revealed that children chose the correct character more than would be expected by chance in the belief condition (86% correct), $z = -5.21$, $p < .001$, and in the probability condition (85% correct), $z = -5.11$, $p < .001$. In the belief condition, seven 5-year-olds and three 6-year-olds initially failed the prompt question. In the probability condition, seven 5-year-olds and four 6-

year-olds initially failed the prompt question. Only one 5-year-old in each condition initially failed the prompt question on both trials. All children answered the control prompt question correctly. These findings demonstrate that children's difficulty when inferring surprise is not a result of an inability to reason about beliefs or probability.

Of primary interest was whether children who were asked the belief and probability prompt questions would perform better than children who were asked the control prompt question. Figure 5 shows children's mean number of correct responses. A GEE binary logistic regression with age (5, 6) and condition (belief, probability, control) as between-subject factors revealed a marginally significant effect of condition, *Wald* $X^2(df = 2, N = 120) = 5.78$, $p = .056$, a marginally significant effect of age, *Wald* $X^2(df = 1, N = 120) = 3.01$, $p = .083$ (6-year-olds chose the correct character more often than 5-year-olds), and a marginally significant condition by age interaction, *Wald* $X^2(df = 2, N = 120) = 5.78$, $p = .056$. Pairwise comparisons revealed that children in the probability prompt condition performed significantly better than children in the control prompt condition, $p = .012$, but no differences were found between children's performance in the belief and control prompt conditions, $p = .796$. Children's performance in the probability prompt condition was also significantly better than their performance in the belief prompt condition, $p = .040$. Next, we explored the marginally significant condition by age interaction.

We explored each age group separately and found that 6-year-olds' performance differed significantly by condition, *Wald* $X^2(df = 2, N = 120) = 7.98$, $p = .019$, but 5-year-olds' performance did not, *Wald* $X^2(df = 2, N = 120) = 0.84$, $p = .656$. Pairwise comparisons revealed that 6-year-olds performed significantly better when asked the probability prompt ($M = 1.80$, $SD = .523$; these means reflect scores summed across the two trials) than when asked the belief prompt (*M* = 1.05, *SD* = 1.00), p =.002, or the control prompt (*M* = 1.20, *SD* = .768), p =.003, but no differences were found between the belief prompt and the control prompt, p = .585. Further pairwise comparisons revealed that 6-year-olds performed significantly better than 5-year-olds when prompted about probability, p =.003, but performed similarly to 5-year-olds when prompted about belief, p = .497, or the control prompt, p =.426. We then summed children's correct responses across both trials for a maximum score of two. Wilcoxon sign tests revealed that only 6-year-olds in the probability prompt condition performed above chance levels, z = -3.77, $p < .001$.

These results reveal that prompting 6-year-olds (but not 5-year-olds) to consider probability when inferring surprise improved their judgments, but prompting children of both ages about beliefs did not. Therefore, Experiment 3 provides two novel insights into children's reasoning about surprise: First, asking children to consider probability appears to be more powerful than asking them to consider beliefs when inferring surprise (at least in this task). Second, children may not spontaneously relate probabilities to surprise when they first appreciate its relevance (around age 6), but they do see the importance of probability when prompted to consider it.

Experiments 1 to 3 used a forced-choice paradigm in which children had to determine which of two characters is more surprised. A concern with this methodology is that children may not have thought that *either* character was surprised, and only chose a character because they were required to do so. Another concern with Experiment 3, is that the control condition might have hindered children's performance because the prompt question was irrelevant to the task. We conducted a final study to address these concerns, while also attempting to replicate the findings from Experiment 3.

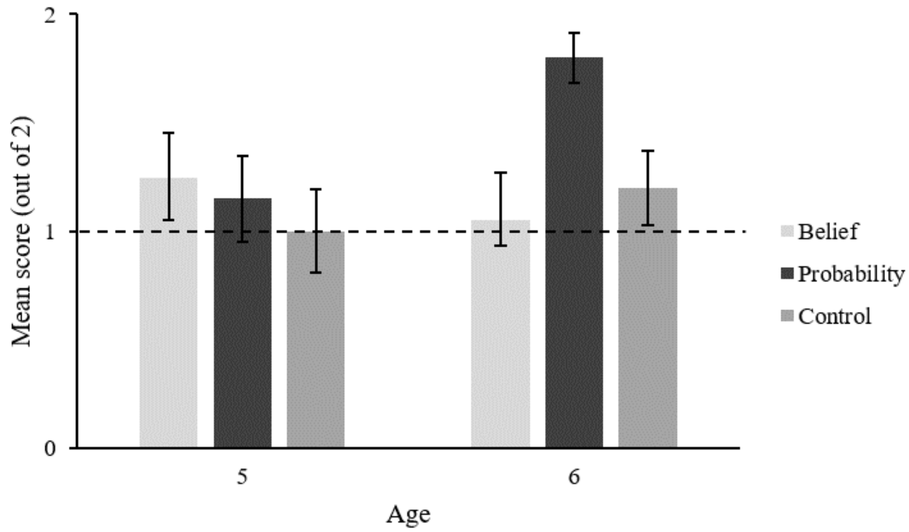


Figure 5. Mean scores for children's surprise judgments in Experiment 3 when asked the belief, probability, and control prompt questions. Error bars show ± 1 standard error of the mean. Dotted line at a score of 1 represents chance performance.

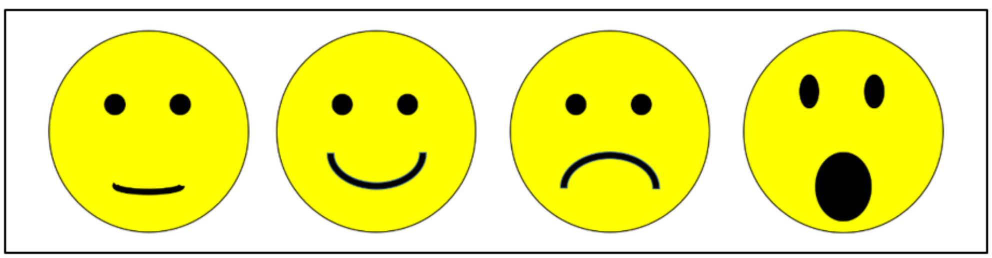
Experiment 4

Experiment 4 further examines the effects of prompting children to consider a character's chances and beliefs of receiving a gumball using a more conservative design. In this experiment, children were shown a single character in front of a gumball machine, and were asked either a belief, probability, or control prompt question. The character then received a minority coloured gumball, and children were asked two questions. The first asked how the character felt about receiving a gumball. We asked this question to allow children to express that the character was happy – in piloting, we found that children were strongly inclined to say this, regardless of what question about emotions was asked. This is unsurprising given that the character received a nice treat. The second question then asked how the character felt after seeing that the gumball was of the minority colour. Of key interest here was whether attributions of surprise would differ depending on the prompt question children were asked. Only 6-year-olds were tested as they were the only age group that showed differences between the prompt conditions in Experiment 3.

Method

Participants. Sixty 6-year-old children participated (*M* = 6;6; range = 6;0 – 6;11; 29 girls). Children were recruited and tested at schools or in a museum.

Materials and procedure. Children were first familiarized to four faces – a neutral face, a happy face, a sad face, and a surprised face. Children were asked to identify each of the faces in a forced-choice pointing task (e.g., "Which face shows feeling happy?"); two different orders were randomly generated and used when asking children to identify the faces and children were randomly assigned to an order. After children identified each face, the experimenter repeated which emotion each face depicted, in the order that the faces were asked about. See Figure 6 for an example of the faces and questions asked.

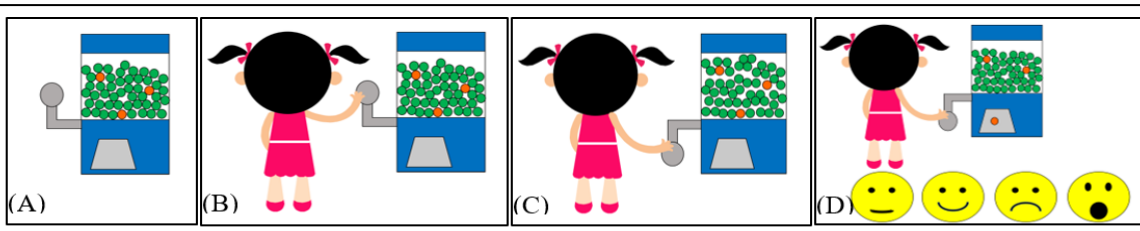


Which face shows feeling surprised? Which face shows feeling happy? Which face shows feeling normal? Which face shows feeling sad? That's right! This face shows feeling surprised. This face shows feeling happy. This face shows feeling normal. And this face shows feeling sad.

Figure 6. Example script and faces used to familiarize children to the four emotions in Experiment 4.

Next, children were told two stories (2 trials, order counterbalanced across participants). In one story, a girl was at a gumball machine with many green gumballs and only a few orange ones; in the other story, a boy was at a gumball machine with many purple gumballs and only a few red ones. For ease of exposition, we describe the procedure from the story with the girl. Children saw a girl appear at a gumball machine, and were told that she wanted a gumball. Children were randomly assigned to one of three prompt conditions. The belief prompt asked, “Which colour gumball does the girl *think* she’s going to get?”, the probability prompt asked, “Which colour gumball does the girl have a *better chance* of getting?”, and the control prompt asked, “Which colour gumball does the girl *see more of*?” After the prompt question, the girl pulled the handle of the machine, and received an orange gumball. The faces showing four emotions then appeared, and children were asked, “How does the girl feel about getting a gumball?”, and “How does the girl feel when she sees that the gumball is orange?”, respectively. Children responded by pointing to one of the faces. See Figure 7 for a sample of the story and script for the probability prompt.

A few children initially failed the prompt question (see Results). When this happened, the experimenter said, “Let’s hear the story again”, repeated the story from the beginning, and re-asked the prompt question. All children answered the prompt question correctly after the repetition.



(A) Here’s a gumball machine. It has green gumballs and it has orange gumballs. (B) And look, here’s a girl. She wants a gumball. *Which color gumball does the girl have a better chance of getting?* (C) To get a gumball, the girl pulls down the handle, and the machine shakes up all the gumballs. (D) Look! She got an orange gumball. So now I have a question for you. *How does the girl feel about getting a gumball? How does the girl feel when she sees that the gumball is orange?*

Figure 7. Sample slides and script of the probability prompt from Experiment 4.

Results and Discussion

We first examined children's *initial* responses to the prompt questions. Wilcoxon sign tests revealed that children chose the correct coloured gumballs more than would be expected by chance in the belief condition (70% correct), $z = -2.53$, $p = .011$, and in the probability condition (85% correct), $z = -3.50$, $p < .001$. In the belief condition, eleven 6-year-olds initially failed the prompt question. In the probability condition, five 6-year-olds initially failed the prompt question. Only one 6-year-old in each condition initially failed the prompt question on both trials. All children answered the control prompt question correctly.

As predicted, when asked how the character felt about receiving a gumball, the majority of children (75%) answered "happy". Of primary interest was whether children who were asked the belief and probability prompt questions would attribute surprise to the character more than children who were asked the control prompt question when asked how the character felt about receiving a minority coloured gumball. Figure 8 shows the mean number of times children indicated that the character would be surprised. A GEE binary logistic regression with condition (belief, probability, control) as a between-subject factor revealed a significant effect of condition, *Wald* $X^2(df = 2, N = 60) = 8.97$, $p = .011$. Pairwise comparisons revealed that children in the probability prompt condition ($M = 1.20$, $SD = .951$), attributed surprise to the character significantly more than children in the control prompt condition ($M = 0.35$, $SD = .671$), $p = .001$, but no differences were found between children's performance in the belief ($M = 0.65$, $SD = .875$) and control prompt conditions, $p = .212$. Children in the probability prompt condition also attributed surprise to the character marginally more than children in the belief prompt condition, $p = .051$.We then summed children's attributions of surprise across both trials. Wilcoxon sign tests revealed that only children in the probability prompt condition chose surprise at above chance levels, with chance being 0.50 out of 2, as there were four possible emotions to choose from and two trials, $z = -2.69$, $p = .007$.

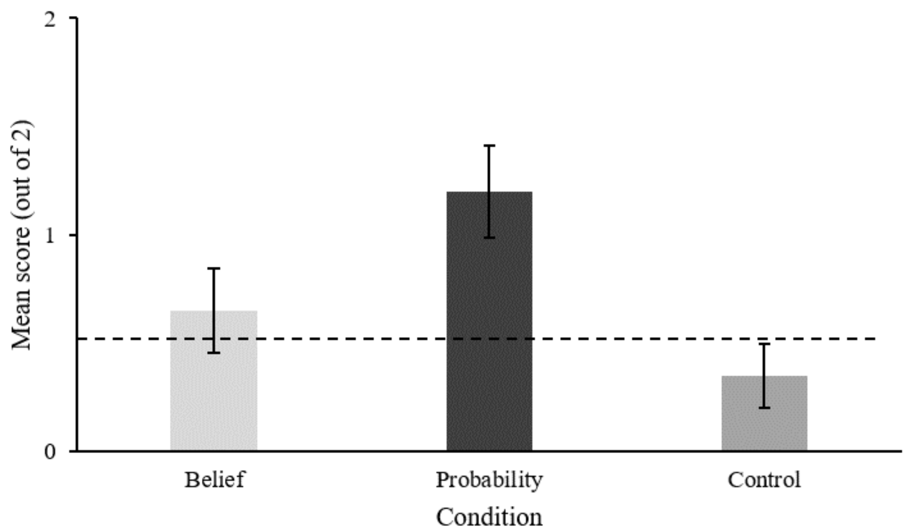


Figure 8. Mean number of times children attributed surprise in Experiment 4 when asked the belief, probability, and control prompt questions. Error bars show ± 1 standard error of the mean. Dotted line at a score of 0.5 represents chance performance.

These results replicate Experiment 3 using a different methodology. We found that prompting 6-year-olds to consider probability led to an increase in their surprise attributions, but prompting them to consider belief did not have this effect. By allowing children to choose between four different emotions, we were able to see the cases in which children would choose surprise over other emotions. We also ruled out the concern that the control prompt in Experiment 3 might have confused children, as they continued to perform poorly in the control prompt condition in this experiment. Together, Experiments 3 and 4 show that for 6-year-olds, the link between probability and surprise is stronger than the link between belief and surprise.

Although all of our Results sections included analyses from both trials, we also ran analyses that only included data from the first trial. The results of these analyses were qualitatively similar with our main analyses, though some significant results became marginal due to the loss of statistical power. Because these analyses do not change the interpretation of our findings, we did not include them.

General Discussion

In four experiments, we examined children's ability to use probability when explicitly inferring other people's surprise. Children's performance improved with age. Whereas 4- and 5-year-olds did not use probability to infer surprise, 6-year-olds did so inconsistently, and only 7-year-olds did so reliably (Experiment 1). When getting a red gumball was impossible, and maximally surprising, children's performance did not improve (Experiment 2). When children were prompted to consider either the characters' chances or beliefs of getting a red gumball, only the prompt about probability improved 6-year-olds' performance (Experiments 3 and 4). Together these findings suggest that children aged 5 and under fail to use probability to infer surprise, and that children aged 6 only have a limited ability to make these inferences (i.e., they require prompting to consistently infer surprise from probability). It is only by age 7 that children have a robust ability to infer surprise from probability.

The present experiments provide the first evidence that children use probability to infer others' emotions. Theories of children's emotional understanding posit two ways in which children understand the causes of emotions – relying on memorized scripts (e.g., Harris et al., 1987; Widen & Russell, 2010, 2011), and considering mental states (e.g., Skerry & Spelke, 2014; Wellman & Woolley, 1990). Neither theory refers to probability or related concepts. Our experiments suggest that these theories are insufficient for characterizing the factors that influence children's emotion attributions, at least in the case of reasoning about surprise. Future work is required to determine whether probabilistic inference influences children's other emotion attributions.

Furthermore, our findings advance knowledge of children's understanding of surprise. They call into question the notion that children mainly conceptualize surprise in terms of beliefs, which is prevalent in the developmental literature, given that most research on surprise in childhood focuses on beliefs (e.g., Hadwin & Perner, 1991; MacLaren & Olson, 1993; Ruffman & Keenan, 1996; Wellman & Banerjee, 1991). We found that 6-year-olds' performance improved when they were prompted to consider probability, but not when they were prompted to consider the character's belief. If surprise were purely belief-based in childhood, then prompting children to consider beliefs should have improved their performance. At a minimum, these findings suggest that children's understanding of surprise is not just belief-based, but also probability-based.

Development and a Conceptual Shift

The development of children's ability to use probability to explicitly infer surprise is strikingly slow given that probability influences children's expectations early in development. Very young children correctly reason about probability (e.g., Denison et al., 2006, 2013; Girotto et al., 2016) and use it to make sophisticated social inferences (Kushnir et al., 2010; Ma & Xu, 2011). In our studies, we also found that children correctly responded to the prompt question about probability (Experiments 3 and 4), demonstrating that they are capable of probabilistic reasoning. Yet our findings suggest that children cannot use their understanding of probability to infer surprise until at least 6 years of age.

If even infants and preschoolers use probability in social inferences, why do young children struggle to use probability when inferring surprise? One possibility is that their difficulty stems from immature inhibitory control. Inhibitory control, the ability to suppress impulsive responses to stimuli, develops over children's preschool years, improving immensely between the ages of 3 and 6 (e.g., Carlson & Moses, 2001; Carlson & Wang, 2007). In our experiments, both machines produce a red gumball; however, one machine always has more red gumballs than the other. Young children may have difficulty inhibiting the impulse to match the outcome (red gumball) to the mostly red gumball machine. However, this account is unlikely to explain all of the difficulties observed by children in our experiments. Older 4-year-olds are proficient at conflict inhibition tasks, which require them to provide a response that is incompatible with their impulsive response (Carlson, 2005; Carlson & Moses, 2001). Thus, 5-year-olds should have easily overcome the inhibitory demands of our task.

It is more plausible that children's difficulty stems from a conceptual deficit. On this view, children aged 5 and under have independent understandings of probability and surprise, but do not see how they relate. By age 6, a conceptual change occurs, in which children come to integrate and relate their understandings of probability and surprise, although it is not until age 7 that they *spontaneously* link the two concepts. Consistent with this account, previous studies show that children correctly reason about probability in explicit tasks by age 4 (e.g., Denison et al., 2006, 2013; Girotto et al., 2016), and they successfully infer surprise at ages 4 – 5 (e.g., Hadwin & Perner, 1991; MacLaren & Olson, 1993; Wellman & Banerjee, 1991; Wellman & Bartsch, 1988). Also, in our third experiment, when 5-year-olds were explicitly prompted to consider probability, they were still unable to use this information to infer surprise, suggesting that they had difficulty integrating probability with surprise. Future research can explore why the connection between probability and surprise arises relatively late in development, and what experiences are needed for this conceptual shift to occur.

Open Questions and Future Directions

Previous studies suggest that children's understanding of surprise is belief-based (e.g., Hadwin & Perner, 1991; MacLaren & Olson, 1993; Wellman & Banerjee, 1991; Wellman & Bartsch, 1988). Further, participants in our experiments correctly anticipated what the characters believed they would get. Therefore, it is perhaps puzzling that our belief prompts did not improve children's surprise judgments. Why was this prompt ineffectual? One possibility is that although children as young as age 4 use beliefs to *explain* surprise (e.g., Wellman & Banerjee, 1991), they might not use beliefs to *infer* surprise until later. The most careful study of children using belief to infer surprise found that they only started making these inferences at age 7, and that younger children instead used ignorance to infer surprise (Ruffman & Keenan, 1996). This could explain why our belief prompts were ineffectual.

Although I only investigated children's ability to use probability to infer surprise, they might also use probability to infer other emotions. For example, anyone would be happy about winning a contest; however, someone who wins a contest that 500 other people entered might be happier and more excited than someone who wins a contest that one other person entered. Similarly, losing a contest that one other person entered would be more disappointing than losing a contest that 500 other people entered – the outcome is the same, but the probability of the outcome occurring changes the degree of emotion experienced. If children do use probability to infer other emotions, this would further suggest that theories of emotion attribution should be expanded to acknowledge the role of probability. Thus, in the next chapter, I investigate whether children use probability to infer other people's happiness.

Chapter Three: Children use probability to infer other people's happiness (Paper Two)

A version of this paper is published:

Doan, T., Friedman, O., & Denison, S. (2020). Young children use probability to infer happiness and the quality of outcomes. *Psychological Science*, *31*(2), 149-159. doi: 10.1177/0956797619895282

Our happiness with an outcome depends on the likelihood of better or worse alternatives. For example, in gambling simulations, adults feel worse about not winning any money if their chances of winning were high, rather than low (Mellers et al., 1997). Such emotional reactions are widely believed to depend on counterfactual comparisons in which people's emotions are intensified when they know they could have had a better or worse outcome (e.g., Bell, 1985; Loomes & Sugden, 1986; Shepperd & McNulty, 2002; van Dijk & van der Pligt, 1997).

People may also consider the probability of better or worse outcomes when inferring others' happiness. If so, this would suggest that people's intuitive theory of happiness is linked with their understanding of probability. Here we explore this proposal from a developmental perspective.

Much previous work suggests that young children have difficulty seeing the link between alternative outcomes and emotions. Children do not experience counterfactual emotions themselves until they are 5-7 (O'Connor et al., 2012; Weisberg & Beck, 2010). Further, they have difficulty using counterfactual comparisons to anticipate others' emotions (e.g., Guttentag & Ferrell, 2004; see Beck & Riggs, 2014 for a review). In one study, 5-7-year-olds considered an agent who could choose one of two sealed boxes and keep its contents (Weisberg & Beck, 2010). The chosen box was revealed to contain just a few stickers, whereas the unchosen box held many more. Children did not indicate that the agent would feel badly about having chosen the box with fewer stickers, even though children did feel badly when they were the ones to make this worse choice for themselves. Similarly, young children often fail to infer that expectations (which provide counterfactual alternatives) can influence happiness (Lara et al., 2019; but see Asaba et al., 2019 for evidence of this ability in children aged 5). Importantly, though, none of these studies specifically manipulated the odds of better or worse outcomes.

Providing young children with information about the probability of better or worse outcomes might allow them to infer happiness. This information may matter because the odds of an alternative outcome affects its impact. For example, children might use probabilistic reasoning to first assess the quality of an outcome (i.e., how good or bad it is), and then use this assessment to infer happiness. When someone with a high probability of winning money does not actually win any, children could first assess that this outcome is bad or unlucky, and then conclude that the person is unhappy. It is plausible that children might make probability-based inferences of happiness because infants and preschoolers make simple probabilistic inferences (e.g., Denison et al., 2006; Denison & Xu, 2010; Téglás et al., 2007), and preschoolers also use probability in social judgments (e.g., inferring preferences; Kushnir et al., 2010; Ma & Xu, 2011).

To investigate whether children infer other people's happiness by considering the probability of better and worse outcomes, we focussed on children aged 4-6. Children in this age range infer happiness by relying on memorized scripts, such as *people are happy when they receive presents* (e.g., Widen & Russell, 2010, 2011; also see Fehr & Russell, 1984). They also make these inferences by considering people's mental states (e.g., Harris et al., 1989; Wellman & Woolley, 1990). For example, 5-year-olds predict a boy will be happy if he believes a box contains candy, even if its actual contents are undesirable (Hadwin & Perner, 1991). Perhaps 4-6-year-olds also infer happiness by considering probability.

To investigate this possibility, 4-6-year-olds in Experiment 1 rated how happy a girl felt upon receiving equal numbers of yummy and yucky gumballs from a gumball machine. Across conditions, the girl either had a high chance of getting mostly yummy ones, or a high chance of getting mostly yucky ones. Experiment 2 added a condition where the girl was equally likely to get yummy or yucky gumballs. Experiments 3 and 4 investigated children's judgments about outcome quality. Experiment 5 investigated adults' judgments about outcome quality and people's happiness.

Experiment 1

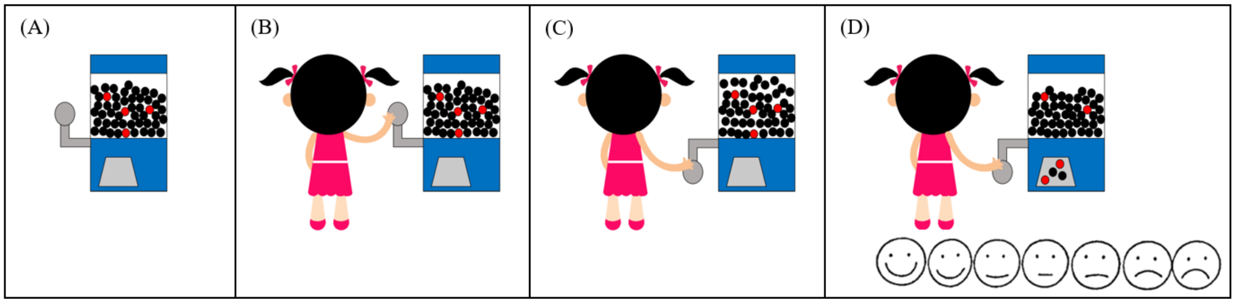
Methods

Participants. We tested 180 children: 60 4-year-olds (*M* = 4;6 [years; months]; range = 4;0 – 4;11; 29 girls), 60 5-year-olds (*M* = 5;5; range = 5;0 – 5;11; 26 girls), and 60 6-year-olds (*M* = 6;5; range = 6;0 – 6;11; 34 girls). Our sample size of 30 participants per cell means that we might have low power to detect small to medium effects, but we followed it based on our lab stopping rule when these data were collected. Our lab chose this stopping rule because we felt it sufficed to reveal effects in previous developmental experiments using similar designs. Children from all experiments were individually tested at schools and daycares in a mid-sized Canadian city in Southwestern Ontario. All data collection occurred between April, 2017 and April, 2019.

Materials and procedure. All materials were shown on a laptop computer. Children were told a story about a gumball machine containing yummy red gumballs and yucky black ones. Children in each age range were randomly assigned to see this story in one of two between-subjects conditions. In the Mostly Yummy condition, the gumball machine contained many red gumballs and just a few black ones (46 red; 4 black). In the Mostly Yucky condition, the gumball machine contained the reverse distribution (46 black; 4 red).

A girl appeared beside the gumball machine in both conditions. She wanted red gumballs and children were asked a comprehension question to confirm that they understood. The girl pulled the handle of the machine and received two red and two black gumballs. Children were asked to rate how she felt using a 7-point happy face scale, ranging from extremely sad to extremely happy. See Figure 9 for a sample of the story and script.

In both conditions, the girl had an extremely high chance of mostly or exclusively receiving the more common type of gumball. For example, in the Mostly Yucky condition, she had a 97% chance of getting at least 3 black gumballs. Hence, getting 2 red and 2 black gumballs is a relatively good outcome in the Mostly Yucky condition, and a relatively bad outcome in the Mostly Yummy condition. Thus, if children infer happiness by considering the probability of better or worse outcomes, they should rate the girl as happier in the Mostly Yucky condition than in the Mostly Yummy condition.



(A) Here's a gumball machine. It has red gumballs and it has black gumballs. The red gumballs are very, very yummy and the black gumballs are very, very yucky. (B) And look, here's a girl. She wants a red gumball. *What color gumball does the girl want?* (C) To get some gumballs, the girl pulls down the handle, and the machine shakes up all the gumballs. (D) Look! She got some gumballs. So now I have a question for you. *How does the girl feel about the gumballs that she got?*

Figure 9. Sample slides and script for the Mostly Yucky condition. The Mostly Yummy condition was identical, but the distribution of yummy and yucky gumballs was reversed.

Results and Discussion

Figure 10 shows children's mean happiness ratings; also see Table 1 for means and standard deviations for all experiments. Data for all experiments are available at https://osf.io/e3a2k/?view_only=505fab0517914db1b065b2285748d565. A 2 (condition: Mostly Yummy, Mostly Yucky) x 3 (age: 4, 5, 6) ANOVA revealed a significant age by condition interaction, $F(2,174) = 4.39$, $p = .014$, $\eta_p^2 = .048$. There were no main effects of condition, $F(1,174) = 3.40$, $p = .067$, $\eta_p^2 = .019$, or of age, $F(2,174) = 2.09$, $p = .127$, $\eta_p^2 = .023$. We explored each age group separately and found that both 5- and 6-year-olds rated the girl as significantly happier in the Mostly Yucky condition than in the Mostly Yummy condition, $t(58) = -2.07$, $p = .043$, $d = 0.53$, and $t(58) = -3.07$, $p = .003$, $d = 0.79$, respectively. However, 4-year-olds did not show differences between the two conditions, $t(58) = 1.18$, $p = .243$, $d = 0.31$.

These findings suggest that 5- and 6-year-olds can use probability to infer happiness However, it is unclear whether this effect was mostly driven by one condition (e.g., children inferring disappointment in the Mostly Yummy condition, without inferring relief in the Mostly Yucky condition).

We examine this in Experiment 2 by seeking to replicate Experiment 1 with 5-6-year-olds, and adding a third baseline condition where the girl was equally likely to get yummy and yucky gumballs.

Table 1

Means and standard deviations for ratings in Experiments 1 – 5

Experiment	Judgment-type	Age	Condition		
			Mostly Yummy	Mostly Yucky	50/50
1	Happiness	4	1.07 (2.32)	0.33 (2.50)	--
		5	0.47 (2.45)	1.63 (1.88)	--
		6	0.83 (1.86)	2.13 (1.38)	--
2	Happiness	5	1.00 (2.20)	1.67 (1.83)	1.30 (2.25)
		6	0.17 (1.74)	1.60 (1.22)	1.33 (1.79)
3	Quality	4	0.40 (2.69)	1.87 (1.74)	--
		5	0.40 (2.33)	1.23 (1.79)	--
		6	0.10 (1.49)	1.60 (1.69)	--
4	Happiness	4	1.05 (2.42)	0.80 (2.42)	--
	Quality	4	0.85 (1.85)	1.95 (1.58)	--
5	Happiness	Adults	-0.14 (1.63)	0.36 (1.64)	--
	Quality	Adults	0.00 (1.58)	0.51 (1.42)	--

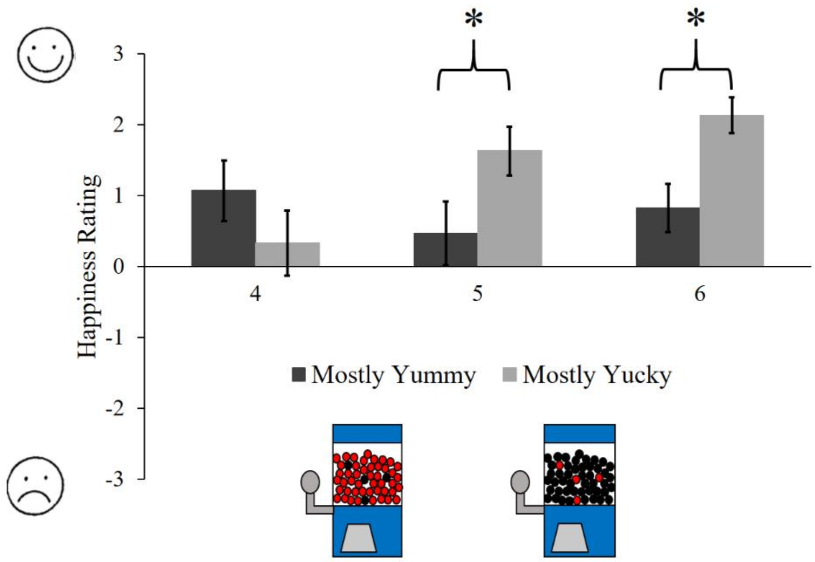


Figure 10. Children's mean happiness ratings, ranging from -3 (extremely sad) to 3 (extremely happy). Error bars show ± 1 standard error of the mean.

Experiment 2

Methods

Preregistration. This study was conducted after Experiments 1 and 3, and was preregistered at http://aspredicted.org/blind.php?x=3ae2dy. The preregistration covered the number of participants, design, and analysis plan.

Participants. We tested 180 children: 90 5-year-olds (*M* = 5;6; range = 5;0 – 5;11; 47 girls), and 90 6-year-olds (*M* = 6;6; range = 6;0 – 6;11; 41 girls). One additional child was tested and excluded because they did not provide a rating.

Materials and procedure. We repeated the procedure from Experiment 1, but added a third 50/50 baseline condition in which the gumball machine contained half yummy and half yucky gumballs (25 red, 25 black). Children in each age range were randomly assigned to one of the three conditions.

Results and Discussion

Figure 11 shows children's mean happiness ratings. A 3 (condition: Mostly Yummy, Mostly Yucky, 50/50) x 2 (age: 5, 6) ANOVA revealed a main effect of condition, $F(2,174) = 4.99$, $p = .008$, $\eta_p^2 = .054$. There was no effect of age, $F(1,174) = 1.08$, $p = .301$, $\eta_p^2 = .006$, and no age by condition interaction, $F(2,174) = 0.97$, $p = .382$, $\eta_p^2 = .011$. Follow-up t-tests revealed that children rated the girl as significantly happier in the Mostly Yucky condition than in the Mostly Yummy condition, $t(118) = -3.21$, $p = .002$, $d = 0.59$, and happier in the 50/50 condition than in the Mostly Yummy condition, $t(118) = -2.00$, $p = .048$, $d = 0.37$. Happiness ratings did not significantly differ between the Mostly Yucky and 50/50 conditions, $t(118) = 0.97$, $p = .335$, $d = 0.17$.

We replicated the findings that 5-6-year-olds use probability to infer happiness. Further, we found this was driven by the understanding that the girl would be relatively upset when she was initially more likely to get a better outcome.

In the remaining experiments, we turn to the finding (from the first experiment) that 4-year-olds did not use probability to infer happiness. Four-year-olds' failure is unlikely to stem from an inability to consider basic probability, as children this age expect a majority item in a distribution to be sampled most often (e.g., Denison et al., 2006; Girotto et al., 2016). One explanation for their failure is that 4-year-olds struggled to give lower happiness ratings for the "good" machine and higher happiness ratings for the "bad" machine. Four-year-olds might find this difficult because it requires inhibiting the impulse to match visually "good" scenes with positive emotions (e.g., Doan et al., 2018).

To examine this possibility, we next investigated whether 4-6-year-olds can use probability to assess the quality of an outcome. As noted above, this could be a necessary step in probability-based inferences of happiness. If the 4-year-olds do use probability to assess quality, this would show that their failure does not stem from inhibitory demands, but may instead reflect a deeper breakdown in the inferences linking probability and happiness.

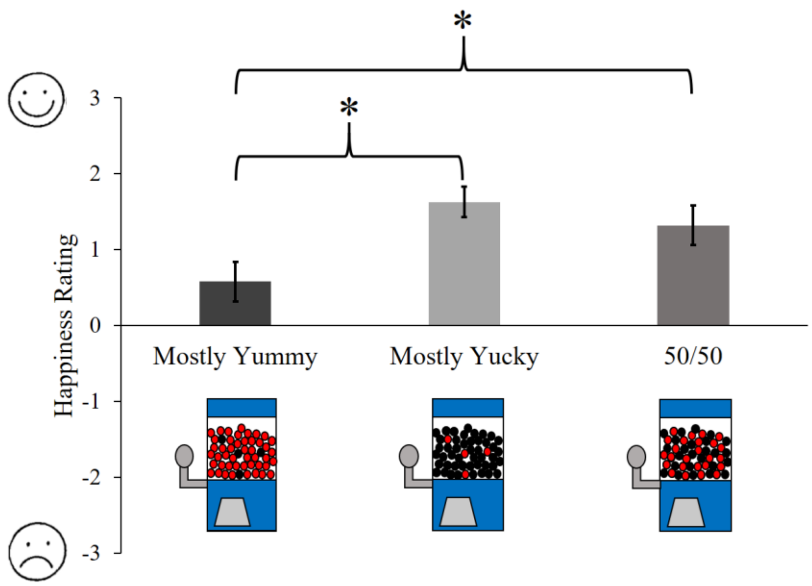


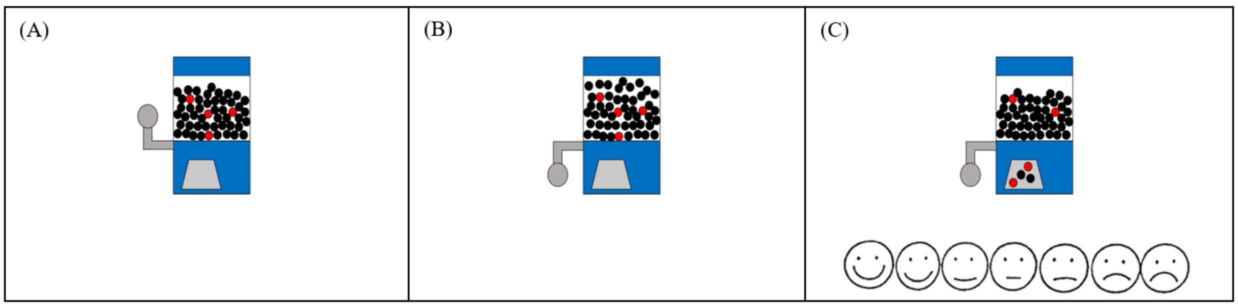
Figure 11. Children's mean happiness ratings, ranging from -3 (extremely sad) to 3 (extremely happy). Error bars show ± 1 standard error of the mean.

Experiment 3

Methods

Participants. We tested 180 children: 60 4-year-olds (*M* = 4;7; range = 4;0 – 4;11; 27 girls), 60 5-year-olds (*M* = 5;5; range = 5;0 – 5;11; 28 girls), and 60 6-year-olds (*M* = 6;5; range = 6;0 – 6;11; 33 girls). Two additional children were tested and excluded because they did not provide a rating.

Materials and procedure. The materials, procedure, and design were identical to Experiment 1, except that the girl was removed from the stories. We used the same rating scale as in the previous experiments, but children were asked to rate how good the outcome was, and not about happiness. Children were randomly assigned to either the Mostly Yummy condition or the Mostly Yucky condition. See Figure 12 for a sample of the story and script for the Mostly Yucky condition.



(A) Here's a gumball machine. It has red gumballs and it has black gumballs. The red gumballs are very, very yummy and the black gumballs are very, very yucky. (B) Let's pull the handle to see what comes out. The machine shakes up all the gumballs. (C) Look what came out! So now I have a question for you. *How good was that?*

Figure 12. Sample slides and script for the Mostly Yucky condition. The Mostly Yummy condition was identical, but the distribution of yummy and yucky gumballs was reversed.

Results and Discussion

Figure 13 shows children's mean quality judgments. A 2 (condition: Mostly Yummy, Mostly Yucky) x 3 (age: 4, 5, 6) ANOVA revealed a main effect of condition, $F(1,174) = 18.07$, $p < .001$, $\eta_p^2 = .094$, where children rated the outcome as better in the Mostly Yucky condition than in the Mostly Yummy condition. There was no effect of age, $F(2,174) = 0.46$, $p = .635$, $\eta_p^2 = .005$, and no age by condition interaction, $F(2,174) = 0.53$, $p = .590$, $\eta_p^2 = .006$.

These findings demonstrate that by age 4, children can determine the relative quality of an outcome, and can inhibit the impulse to match overwhelmingly "good" visual scenes with good ratings. Together, the experiments so far suggest that 4-year-olds can use probability to assess the relative quality of an outcome, but not to infer happiness. However, we must be cautious about this conclusion because it rests on the results of separate experiments. To provide more certainty, we investigated the potential developmental gap between 4-year-olds' assessments of quality and inferences of happiness within a single experiment.

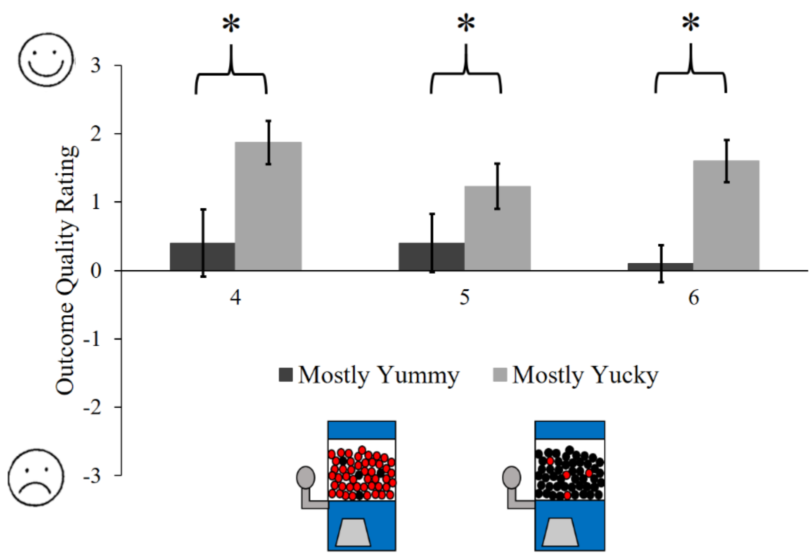


Figure 13. Children's mean outcome quality judgments, ranging from -3 (extremely bad) to 3 (extremely good). Error bars show ± 1 standard error of the mean.

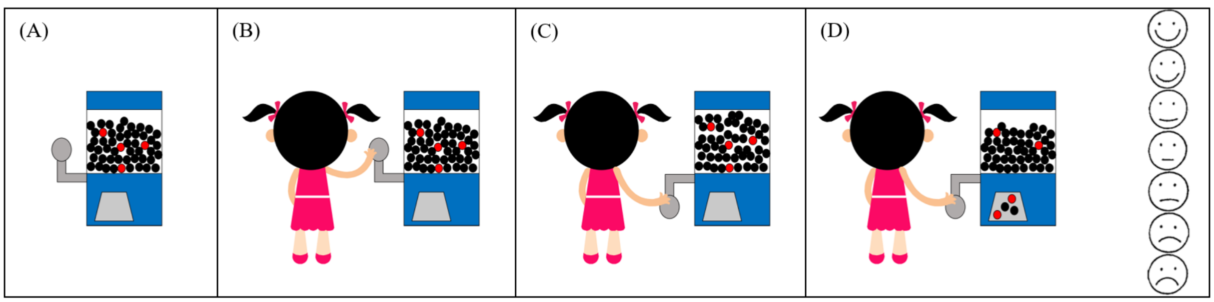
Experiment 4

Methods

Preregistration. This study was preregistered at http://aspredicted.org/blind.php?x=e43tu8. The preregistration covered the number of participants, design, and analysis plan.

Participants. We tested 80 4-year-olds (*M* = 4;6; range = 4;0 – 4;11; 36 girls). Two additional children were tested and excluded because they did not provide a rating. Our preregistered sample size of 40 participants per cell is larger than the stopping rule in the previous experiments because this experiment focussed on just one age group (i.e., meaning that the overall sample of participants was smaller than in the previous experiments).

Materials and procedure. We used a 2X2 design in which condition (Mostly Yummy, Mostly Yucky) was manipulated within-subjects (order counterbalanced across children), and judgment-type (Happiness, Quality) was manipulated between-subjects. In the stories, a girl approached a gumball machine and received 2 yummy gumballs and 2 yucky ones. A happy face scale then appeared, and children were either asked about the girl's happiness ("How does the girl feel about the gumballs that she got?") or about the quality of the outcome ("How good was that?"). See Figure 14 for a sample of the story and script for the Mostly Yucky condition.



(A) Here's a gumball machine. It has red gumballs and it has black gumballs. The red gumballs are very, very yummy and the black gumballs are very, very yucky. (B) And look, here's a girl. (C) To get some gumballs, the girl pulls down the handle, and the machine shakes up all the gumballs. (D) Look! She got some gumballs. So now I have a question for you. *How does the girl feel about the gumballs that she got?* (Happiness question) OR *How good was that?* (Quality question)

Figure 14. Sample slides and script for the Mostly Yucky condition. The Mostly Yummy condition was identical, but the distribution of yummy and yucky gumballs was reversed.

Results and Discussion

Figure 15 shows children's mean happiness and quality ratings. A 2 (condition: Mostly Yummy, Mostly Yucky) x 2 (judgment-type: Happiness, Quality) mixed ANOVA revealed a condition by judgment-type interaction, $F(1,78) = 5.33$, $p = .024$, $\eta_p^2 = .064$. There were no main effects of condition, $F(1,78) = 2.11$, $p = .150$, $\eta_p^2 = .026$, or judgment-type, $F(1,78) = 1.67$, $p = .200$, $\eta_p^2 = .021$.

We conducted follow-up t-tests to better understand the interaction. When rating quality, children gave higher ratings in the Mostly Yucky condition than in the Mostly Yummy condition, $t(39) = -2.83$, $p = .007$, $d = 0.46$. However, when judging happiness, there was no significant difference between the Mostly Yummy and Mostly Yucky conditions, $t(39) = 0.57$, $p = .570$, $d = 0.09$, replicating our previous findings.

The findings replicated the developmental gap in which 4-year-olds succeed in using probability to assess quality, but not to infer happiness. We wondered whether this gap between quality and happiness continues later in the lifespan. If so, this might suggest that probability is linked more closely with quality than with happiness. Our earlier experiments provided some evidence against this possibility. Specifically, they showed that older children use probability to infer both happiness (Experiments 1 and 2) and quality (Experiment 3). But because these were assessed in different studies (which also used slightly different methods), those experiments only provide a weak basis for comparing these judgments. To better examine this possibility, we conducted an online version of this study with adults.

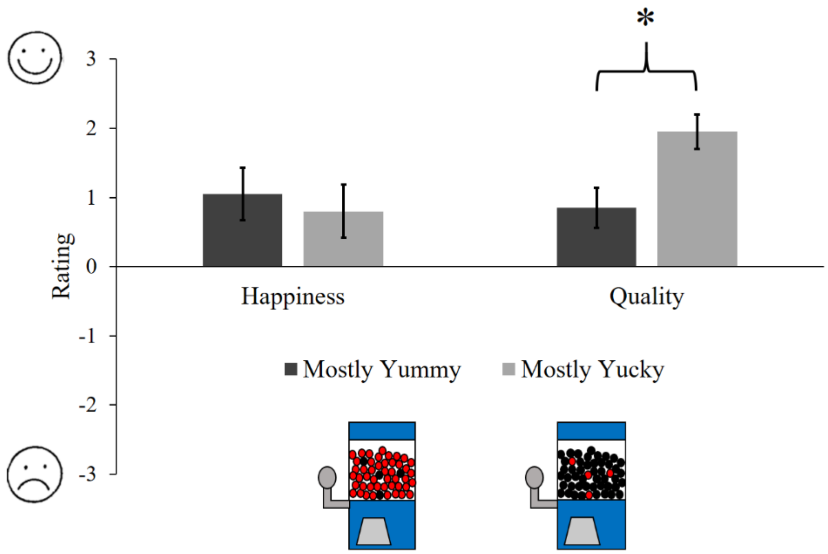


Figure 15. Children's mean happiness and outcome quality ratings, ranging from -3 (extremely sad/bad) to 3 (extremely happy/good). Error bars show ± 1 standard error of the mean.

Experiment 5

Methods

Preregistration. This study was preregistered at http://aspredicted.org/blind.php?x=mt7mz9. The preregistration covered the number of participants, design, and analysis plan.

Participants. The experiment was successfully completed by 254 adults (*M* = 35; 111 females) recruited using Amazon Mechanical Turk. Based on our preregistration, we originally recruited 300 participants, but excluded participants who failed at least one of our two comprehension check questions (*N* = 45). One additional participant was excluded because they did not provide a rating on one trial, and so their data could not be analyzed. Our decision to recruit 300 participants was based on the assumption that would leave us with at least 100 participants per between-subject condition (i.e., after exclusions for failures of comprehension checks). We felt that 100 participants per condition would suffice to reveal effects given findings from previous online studies of adults that used similar designs.

Materials and procedure. The experiment used the same 2X2 design as Experiment 4: condition (Mostly Yummy, Mostly Yucky) was manipulated within-subjects, and judgment-type (Happiness, Quality) was manipulated between-subjects. The stories were similar to those in Experiment 4, except each was conveyed using two images, each shown on a separate page. The first page showed a girl at a gumball machine, with the script, "Here's a girl at a gumball machine. There are red gumballs and black gumballs. The red gumballs are very very yummy and the black gumballs are very very yucky". The second page showed the girl had received two yummy and two yucky gumballs, with the script, "The girl pulled the handle and got some gumballs". The test question ("How does the girl feel about the gumballs that she got?" or "How good was that?") appeared lower down on the page, along with the same happy face scale used with children. Participants responded by clicking on one of the faces.

Results and Discussion

Figure 16 shows adults' mean happiness and quality ratings. A 2 (condition: Mostly Yummy, Mostly Yucky) x 2 (judgment-type: Happiness, Quality) mixed ANOVA revealed a main effect of condition, $F(1,252) = 11.95$, $p = .001$, $\eta_p^2 = .045$, where adults had higher ratings in the Mostly Yucky condition than in the Mostly Yummy condition. There was no main effect of judgment-type, $F(1,252) = 1.22$, $p = .270$, $\eta_p^2 = .005$, and no condition by judgment-type interaction, $F(1,252) = 0.001$, $p = .972$, $\eta_p^2 = .000$.

These findings suggest that for the stories we showed children, adults use probability to a similar extent when inferring happiness and outcome quality. As such, the findings do not support the conjecture that probability is linked more closely with quality than with happiness.

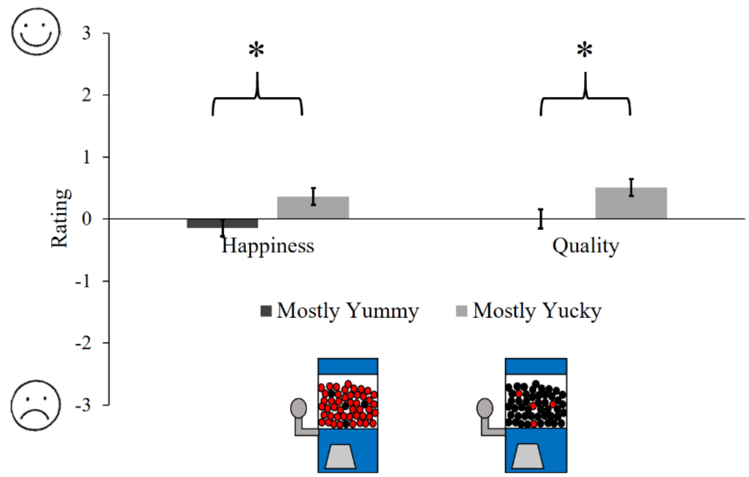


Figure 16. Adults' mean happiness and outcome quality ratings, ranging from -3 (extremely sad/bad) to 3 (extremely happy/good). Error bars show ± 1 standard error of the mean.

General Discussion

We examined whether children and adults use probability to infer people's happiness and the quality of outcomes. For children aged 5 and 6 (and for adults), happiness and quality judgments both depended on whether a better outcome was initially more likely. At age 4, children did not use probability to infer the girl's happiness, but did use it to infer the relative quality of the outcome.

We found that children use probability to infer negative emotions (e.g., disappointment), but not positive ones (e.g., relief, elation). In the second experiment, we found that relative to the baseline condition (equal odds of receiving yummy and yucky gumballs), children gave more negative emotional ratings when a better outcome was initially more likely. But they did not give significantly more positive ratings when a worse outcome was initially more likely. This asymmetry is broadly consistent with many previous findings, including those suggesting that: probability more strongly affects adults' negative than positive emotions (e.g., van Dijk & van der Pligt, 1997); children experience the negative emotion regret before they experience the positive emotion relief (Weisberg & Beck, 2010); and children and adults react more strongly to having less than others (disadvantageous inequality) than to having more (advantageous inequality; e.g., Blake & McAuliffe, 2011; Boyce et al., 2010; Lobue et al., 2011).

We also found that children use probability to assess quality before they can use it to infer happiness. This developmental gap is counterintuitive. If 4-year-olds use probability to assess quality, and likewise understand that negative events lead to unhappiness (e.g., Widen & Russell, 2011), why do they struggle to see how probability can affect emotions? One explanation is that 4-year-olds do not *spontaneously* draw both inferences in this chain (i.e., probability → quality, quality → happiness). When asked how the girl feels, 4-year-olds might not have realized that they should start by inferring the quality of the outcome. On this account, 4-year-olds might do better if prompted to first assess quality. However, alternative explanations for the developmental gap between inferences of quality and happiness are possible. Perhaps older children use probability to infer happiness, without first assessing quality (e.g., by considering probability in relation to desires). On this account, 4-year-olds might fail because they have not connected the disparate conceptual domains of emotion and probability.

Regardless, these accounts leave open deeper questions about how children and adults use information about probability to infer happiness and assess quality. They could use this information to establish a "standard of comparison" against which actual outcomes are considered (e.g., Cadotte, Woodruff, & Jenkins, 1987). For instance, when a good outcome is highly probable, this establishes a high standard of comparison. Children (and adults) may then judge that a mediocre outcome is bad because it falls short of this standard (e.g., a gumball machine with 90% good gumballs sets this as the standard, and so the outcome of 50% good gumballs is disappointing). This type of reasoning might be easier for younger children as the actual outcome can be compared to a standard without specifically considering that a better or worse outcome could have occurred.

Alternatively, children's and adults' inferences could depend on counterfactual comparisons. For example, when the machine mostly contained yummy gumballs, participants may have judged the outcome was relatively bad because they recognized that a better outcome could easily have occurred. At first thought, reasoning in this way seems unlikely for children. Although young children can reason about counterfactuals (Nyhout & Ganea, 2019; also see German & Nichols, 2003; Harris et al., 1996), they may only come to use counterfactuals to infer others' happiness at age 7 (e.g., Guttentag & Ferrell, 2004). However, there is reason to think that the manipulation of probability may have made using counterfactuals to infer emotions easier. When the prior probability of an alternative outcome occurring is high, it may feel closer to happening, thus possibly intensifying the emotion felt when it does not happen. If children considered the *closeness* of the counterfactual alternative outcomes, they may have used it to infer emotions. I look at this possibility more closely in the next chapter.

We next consider two broader implications of the findings. First, our findings advance knowledge of the ways in which children attribute happiness. They show that children do not only infer happiness by relying on memorized scripts (e.g., Russell 1990; Widen & Russell, 2010, 2011), or by considering others' mental states (e.g., Hadwin & Perner, 1991; Wellman & Bartsch, 1988). In addition to these methods for inferring happiness, children also draw on their understanding of probability. Second, our findings advance knowledge of how children use probability information to make social inferences. Preschool-aged children use probability to infer preferences (Kushnir et al., 2010; Ma & Xu, 2011), and older children use it to infer others' surprise (Doan et al., 2018). We extend these findings by showing that 4- and 5-year-olds use probability to infer outcome quality and others' happiness. We found evidence for a developmental lag between children's probability-based inferences of outcome quality and happiness. With development, probabilistic understanding may be gradually integrated into children's understanding of the social world.

Chapter Four: Children use probability and (maybe) close counterfactual reasoning to infer other people's happiness (Paper Three)

A version of this paper is under review:

Doan, T., Friedman, O., & Denison, S. (under review). Oh…so close! Children's close counterfactual reasoning and emotion inferences. *Developmental Psychology.*

People's emotional reactions depend not only on actual events, but also on close counterfactuals – events that did not happen but easily could have (Kahneman & Varey, 1990). Consider the classic scenario from Kahneman and Tversky (1982), in which Mr. Tees missed his flight by five minutes, while Mr. Crane missed his flight by 30 minutes. Who is more upset? Most people agree that Mr. Tees is more upset. This is presumably because he was *closer* to making his flight, making it much easier to imagine a world in which he did. Besides affecting judgments about others' emotional reactions (e.g., Johnson, 1986; Kahneman & Tversky, 1982; Medvec & Savitsky, 1997), close counterfactuals also affect emotional reactions firsthand (e.g., Markman & Tetlock, 2000; McMullen & Markman, 2002; also see Medvec et al., 1995, for evidence of this in Olympic medalists' emotions as rated by others).

Many factors can affect the perceived closeness of a counterfactual outcome. In the example of Mr. Tees and Mr. Crane, closeness depended on temporal proximity – it feels like Mr. Tees could have easily made his flight because he was barely late. Counterfactual closeness can also depend on spatial and numerical proximity (e.g., Ong et al., 2015; Teigen, 1996; Turnbull, 1981). In one study, adults were told about two characters who each spun wheels that were equally divided into alternating sectors of blue, red, and yellow. One wheel had three sectors, while the other had 18 sectors. Although both characters landed on the winning color (red), the character who spun the wheel with many sectors was viewed as luckier, presumably because this character was spatially closer to not winning (i.e., with many sectors, losing colors were spatially proximate; Teigen, 1996).

Counterfactual closeness also depends on probability (e.g., Kahneman & Varey, 1990; Roese & Olson, 2014). If the prior probability of an event is high, then even though it does not occur, people may judge that it nearly did. This in turn, can affect emotions (e.g., Bell, 1985). In a gambling task, people felt worse about winning nothing if their initial chances of winning money were high compared to low, presumably because they were closer to winning in the former case (Mellers et al., 1997; van Dijk & van der Pligt, 1997).

Here we explore children's ability to consider close counterfactual alternatives when inferring other people's emotions. Children can reason counterfactually by age 4 (e.g., Beck et al., 2006; German & Nichols, 2003; Harris et al., 1996; but also see Beck et al., 2010). For example, children learned that a box lights up if one block is placed on it, but not if a different block is used. When the box lit up after both blocks were put on it, 4- and 5-year-olds reasoned counterfactually about whether it would have lit up if each block had not been placed on it (Nyhout & Ganea, 2019). Children can also consider *close* counterfactuals when the actual outcome and the counterfactual alternative are spatially close. In one study, children watched as two toy horses galloped along a table; one horse fell and the other stopped just at the edge. Around age 5, children were able to say which of the two horses almost fell (Beck & Guthrie, 2011).

Children's understanding of the emotional consequences of counterfactuals may develop somewhat later. Children do not experience counterfactual emotions until the ages of 5 to 7 (e.g., Guerini et al., in press; O'Connor et al., 2012; Weisberg & Beck, 2010; but see Weisberg & Beck, 2012, for evidence of slightly earlier development), and they do not use counterfactual comparisons to infer other people's emotions until at least age 7 (e.g., Amsel & Smalley, 2000; Beck & Crilly, 2009; Guttentag & Ferrell, 2004; see Beck & Riggs, 2014, for a review). For example, in Weisberg & Beck (2010), a character won a mundane prize when choosing between one of two opaque boxes. When it was revealed that the character could have won a better prize (a counterfactual possibility), even 7-year-olds did not predict that he would now be sad about the prize he got. This finding suggests that children do not consider counterfactual alternatives when inferring others' emotions.

However, no studies have examined whether children consider the *closeness* of counterfactual alternatives when inferring emotions. For example, in Weisberg and Beck (2010), the character was not closer to winning one prize over the other. Without comparison conditions that manipulate the closeness or the odds of the counterfactual alternative, we cannot know whether children's emotion inferences depend on close counterfactuals. As we saw with Mr. Tees and Mr. Crane, the closeness of the counterfactual alternatives is critical: a disappointing outcome may be especially disappointing if a desirable outcome nearly happened. If children use close counterfactuals to infer emotions, this will provide further support that counterfactual reasoning is central to children's thinking from early in development (Weisberg & Gopnik, 2013).

One recent finding suggests that young children might consider close counterfactuals when inferring emotions. Children saw two bowlers, one whose ball was on a trajectory to knock down many pins and another whose ball was on a trajectory to knock down none. In the end, both knocked down 3 of 6 pins. Children aged 5 expected the bowler who almost knocked-down all the pins to feel worse than the one who almost knocked-down none (Asaba et al., 2019). Children might have based this inference on close counterfactuals – they might have reasoned that the first bowler was more upset because they nearly knocked down all the pins. But as Asaba et al. (2019) suggest, children might have based their inferences on the characters' expectations. Further, children were not asked about their counterfactual thoughts, making it difficult to know whether they were using close counterfactuals to infer emotions.

We conducted three experiments to examine whether children consider close counterfactuals when inferring others' emotions. In our first experiment, we tested children aged 4 to 6 as this is the age range in which they begin to reason counterfactually (e.g., Nyhout & Ganea, 2019). However, our later experiments were limited to children aged 5 to 6 as our youngest children showed difficulty in our tasks. In each experiment, children heard stories about a girl who won a mundane prize. Across conditions, we manipulated whether the girl later discovered that her odds of winning a more attractive prize had been high or low. When the odds were very high, the closest counterfactual alternative was winning the attractive prize. However, when the odds were low, the closest counterfactual alternative was winning a mundane prize. Thus, the closeness of the counterfactual alternative was manipulated by differences in probability across conditions. If children consider close counterfactuals, they should recognize that the girl would be happier when she had a low chance of winning an attractive prize than when she had a high chance. In later experiments, we added a second cue to counterfactual closeness – spatial proximity – the girl was in closer proximity to a more attractive prize in the condition in which it was more probable. Children were also asked a question assessing whether they recognized the closeness of the counterfactual alternative where the girl won a more attractive prize.

In our final experiment, we tested adults. Some studies investigating the influence of close counterfactuals on emotions have confounded close counterfactuals with prior beliefs. For example, in gambling tasks, the odds of winning money are known beforehand, so emotion ratings could have been based on participants' prior beliefs about winning instead of whether they were close to winning (e.g., Mellers et al., 1997; van Dijk & van der Pligt, 1997; but for an exception see Ong et al., 2015). We tested adults to address this concern and to provide a reference point for comparison with children's emotion and close counterfactual responses.

Experiment 1

Children heard a story about a girl who chose one of ten closed boxes and won a regular balloon. She then either learned that most of the other boxes contained special balloons or that most contained regular balloons. Children were asked how the girl felt about getting the regular balloon and whether the girl could have easily gotten a special balloon.

Methods

Participants. We tested 180 4-6-year-olds (*M* = 5;6 [years; months]; range = 4;0 – 6;11; 86 girls). There were 60 4-year-olds (*M* = 4;5; range = 4;0 – 4;11; 29 girls), 60 5-year-olds (*M* = 5;6; range = 5;0 – 5;11; 32 girls), and 60 6-year-olds (*M* = 6;6; range = 6;0 – 6;11; 25 girls), with equal numbers of children per age-in-years randomly assigned to each between-subjects condition. Sample sizes for all experiments were decided in advance based on a stopping rule. Our lab chose the stopping rule of 30 participants per cell because it sufficed to reveal effects in previous developmental experiments using similar designs (e.g., Asaba et al., 2019; Doan et al., 2020; Shaw & Olson, 2015). In all experiments, children were individually tested at schools and daycares in a mid-sized Canadian city in Southwestern Ontario. Demographic information was not formally collected, but the region is predominantly middle-class, and approximately 79% of residents in this region are White, with Chinese and South Asians as the most visible minorities.

This research, submitted under the name, “Social Understanding in Children” (ORE#30395), received ethics clearance through the University of Waterloo’s Research Ethics Committee.

Materials and procedure. All materials were shown on a laptop computer. Children were told a story about a girl who could win either a regular or a special balloon. See Figure 17 for the story and script. The girl stood before ten closed boxes that each contained a balloon (the balloons were not visible). The girl could choose one of the boxes and win the balloon in it. She chose a box and received a regular balloon. Next, she was shown the balloons in the remaining boxes. Children were asked how the girl felt about getting the regular balloon, and responded using a 7-point happy face scale, ranging from extremely happy to extremely sad.

After children rated the girl’s happiness, they were asked whether she could have easily gotten a special balloon, with their responses recorded as either “yes” or “no”. We asked the close counterfactual question in this way as this phrasing has been used to assess counterfactual reasoning with adults in other studies (e.g., Teigen 1995, 1997).

Children saw the story in one of two between-subjects conditions. In the Mostly Special condition, there were eight special balloons and two regular balloons (80% chance of getting a special balloon). In the Mostly Regular condition, there were eight regular balloons and two special balloons (20% chance of getting a special balloon). The scripts were identical across conditions.

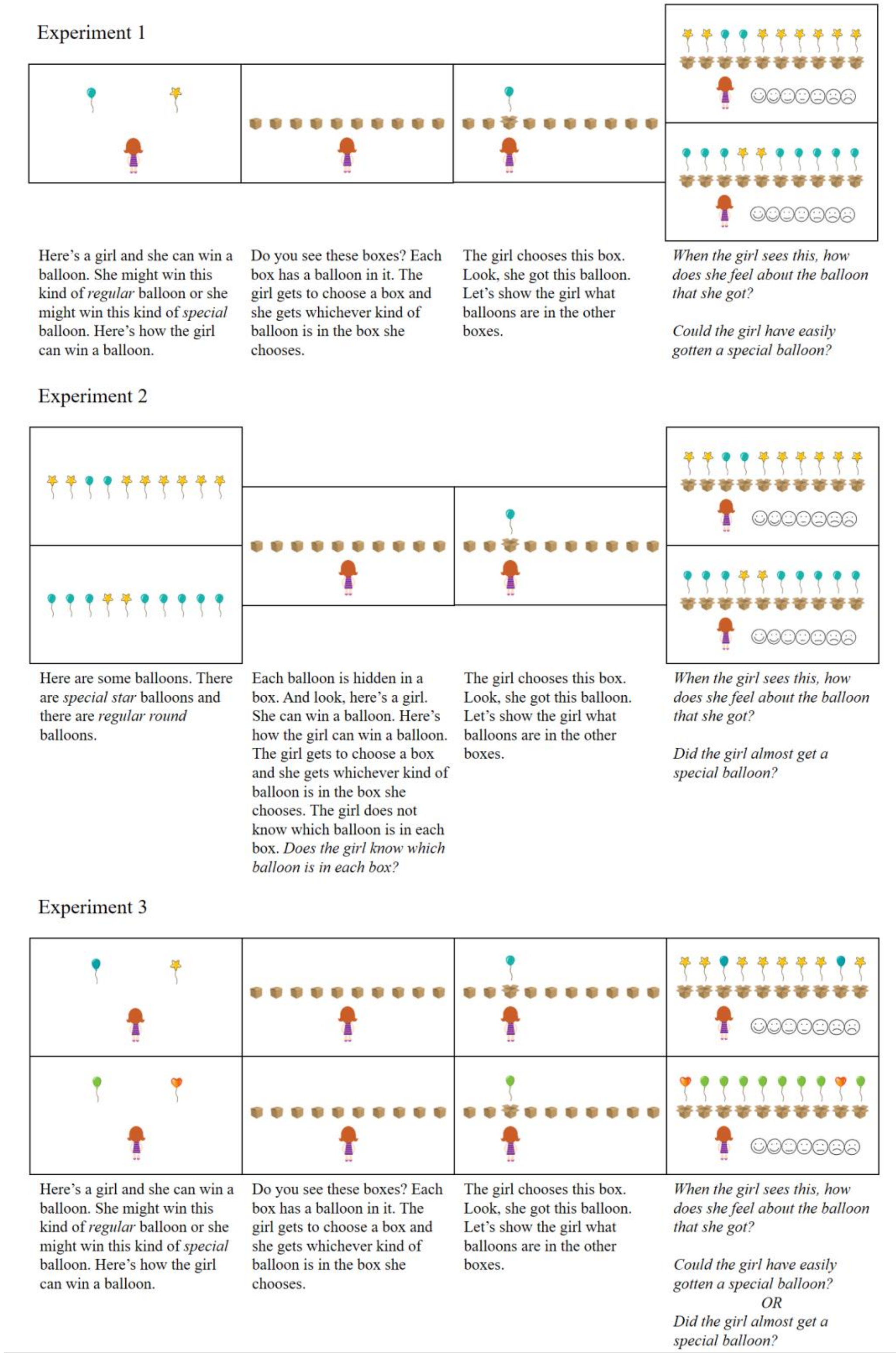


Figure 17. Slides and scripts for Experiments 1-3. The star and heart balloons are the special balloons and the round balloons are the regular balloons.

Results and Discussion

The analyses in all experiments used Generalized Estimating Equations models (GEEs); ordinal logistic for happiness ratings and binary logistic for counterfactual questions. In experiments on children, age in months was mean-centred and entered as a covariate. Table 2 lists the factors entered into each model and all significant and marginal effects.

We first looked at children's happiness ratings. There was a significant main effect of age, *Wald* $X^2(1) = 19.51$, $p < .001$, no main effect of distribution, *Wald* $X^2(1) = 1.88$, $p = .170$, and a significant age-by-distribution interaction, *Wald* $X^2(1) = 12.18$, $p < .001$; see Figure 18. To follow up on this interaction, we explored each age group separately and found that 6-year-olds rated the girl as significantly happier in the Mostly Regular condition than in the Mostly Special condition, *Wald* $X^2(1) = 13.34$, $p < .001$. Five-year-olds did not show differences between the two conditions, *Wald* $X^2(1) = 1.51$, $p = .218$, and 4-year-olds rated the girl as significantly happier in the Mostly Special condition than in the Mostly Regular condition, *Wald* $X^2(1) = 4.51$, $p = .034$. We also explored each condition separately. In the Mostly Special condition, with age, children rated the girl as less happy to receive a regular balloon, *Wald* $X^2(1) = 26.88$, $p < .001$; in the Mostly Regular condition, age did not affect their happiness ratings, *Wald* $X^2(1) = 0.63$, $p = .429$.

Next, we looked at children's responses to the counterfactual question. There was a significant main effect of age, *Wald* $X^2(1) = 4.05$, $p = .044$, no main effect of distribution, *Wald* $X^2(1) = 0.62$, $p = .431$, and a significant age-by-distribution interaction, *Wald* $X^2(1) = 4.09$, $p = .043$; see Figure 18. We explored each age group separately and found that 4-, 5-, and 6-year-olds' responses did not significantly differ between the two conditions, *Wald* $X^2(1) = 0.34$, $p = .558$, *Wald* $X^2(1) = 0.45$, $p = .503$, *Wald* $X^2(1) = 2.10$, $p = .147$, respectively. We also explored each condition separately, and found that in the Mostly Regular condition, with age, children were less likely to say the girl could have easily gotten a special balloon, *Wald* $X^2(1) = 7.69$, $p = .006$; in the Mostly Special condition, age did not affect their judgment, *Wald* $X^2(1) = 0.00$, $p = .993$.

At age 4, children inferred that the girl would be happier when she learns that most of the remaining boxes contained special balloons. Although seemingly counterintuitive, our youngest children may have inferred that the girl is happy because there were overall many special balloons – more special balloons means more happiness. Previous studies have reported similar findings in older children aged 5 and 7 (Guttentag & Ferrell, 2004; McCloy & Strange, 2007). Our older children did not rely on this strategy for their emotion ratings, and by age 6, children inferred the girl would feel sadder about winning a regular balloon when she could have easily won a more attractive one (i.e., compared with when this counterfactual alternative was not close). This finding suggests that children might use counterfactuals to infer others' emotions earlier than previously thought.

However, when directly asked the counterfactual question, even 6-year-olds were insensitive to the difference in closeness between conditions. Perhaps children struggled with our specific question. We asked children whether the girl "could have easily" won a special balloon. In earlier work showing that 5-year-olds can consider close counterfactuals, children were instead asked about whether a counterfactual alternative "almost" happened (Beck & Guthrie, 2011), which may be easier for them to understand. So in the next experiment, we adopted this wording for our counterfactual question.

We also wondered if younger children would succeed if the chronology of the story was changed. Information about the initial odds of winning a special balloon was only provided after it already became evident that this outcome did not occur to reduce the likelihood that children could rely on prior beliefs in their emotion judgements. However, this is a somewhat unusual sequence of events and young children might find it difficult to follow. We examined these possibilities in Experiment 2. That is, we changed the close counterfactual question to be more similar to previous child studies and changed the chronology of the story.

Table 2

Factors and Effects from each GEE Analysis

Experiment	Question	Factors	Effects	Wald X^2	*df*	*p*
1	happiness	distribution	age	19.51	1	<.001
			age*distribution	12.18	1	<.001
	counterfactual	distribution	age	4.05	1	.044
			age*distribution	4.09	1	.043
2	happiness	distribution	age	12.45	1	<.001
			distribution	2.93	1	.087
	counterfactual	distribution	age	4.55	1	.033
			age*distribution	6.83	1	.009
3	happiness	distribution	distribution	12.23	1	<.001
	counterfactual	distribution	distribution	14.92	1	<.001
			age*distribution	7.59	1	.006
		question-type	question-type	13.82	1	<.001
			age*question-type	9.92	1	.002
4	happiness	distribution	distribution	72.84	1	<.001
	counterfactual	distribution	distribution	177.77	1	<.001
		question-type	question-type	3.16	1	.076

Note. In all experiments, distribution was either Mostly Special or Mostly Regular; in Experiments 3 and 4, question-type was either “could the girl have easily gotten” or “did the girl almost get” a special balloon.

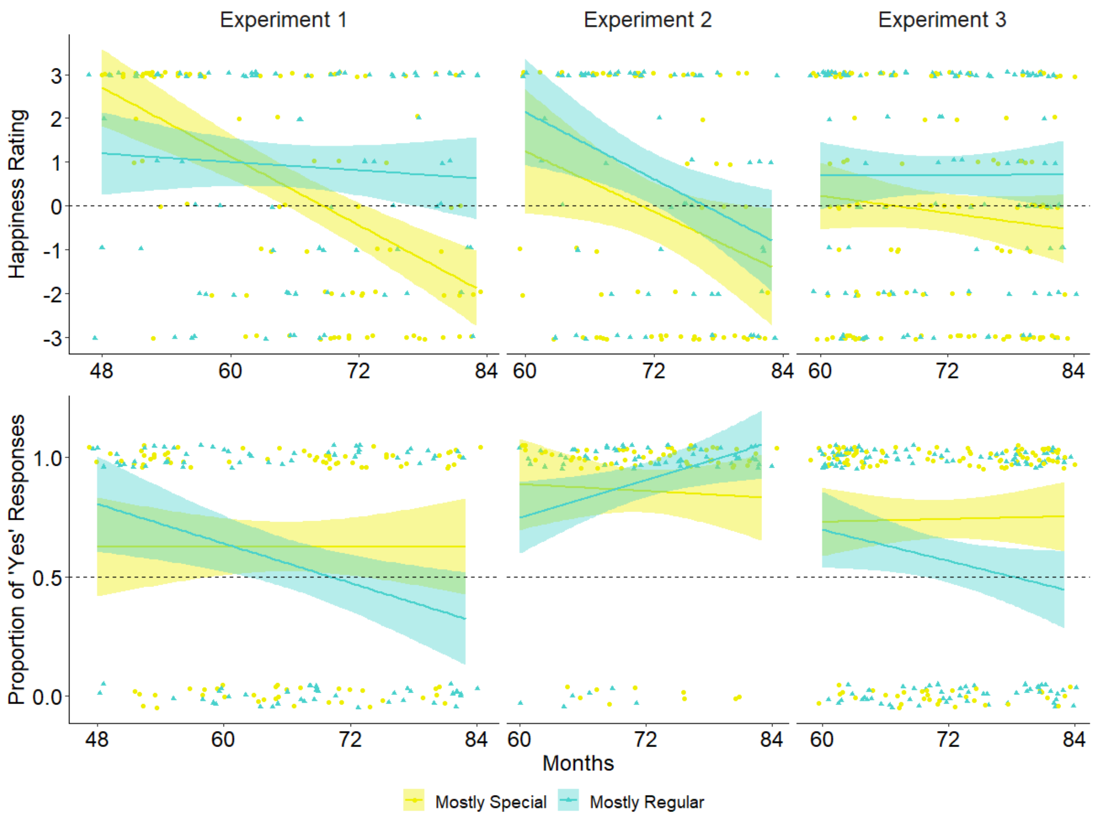


Figure 18. Children's happiness ratings (top row) and counterfactual responses (bottom row) in Experiments 1-3. Happiness ratings ranged from -3 (extremely sad) to 3 (extremely happy). Colored bands show 95% confidence intervals; points are jittered to avoid over-plotting.

Experiment 2

The story was similar to Experiment 1, except children saw the array of balloons at the beginning and were told that the girl did not know which balloons were in each box. Also, instead of asking whether the girl *could have easily* gotten a special balloon, we asked whether the girl *almost* got a special balloon. Lastly, because only the oldest children correctly inferred the girl's happiness in Experiment 1, we tested only 5- and 6-year-olds in the following experiments.

Methods

Participants. We tested 125 5-6-year-olds (*M* = 6;0; range = 5;0 – 6;11; 67 girls); 60 5-year-olds (*M* = 5;6; range = 5;0 – 5;11; 39 girls), and 65 6-year-olds (*M* = 6;5; range = 6;0 – 6;11; 28 girls). Four additional children were tested but excluded for failing the comprehension question three times (*n* = 3) or for having already completed Experiment 1 (*n* = 1). We aimed to test 60 children per between-subjects condition, with equal numbers of children per age-in-years randomly assigned to each condition. However, we accidentally tested five extra 6-year-olds. Thus, we ended up with 63 children in the Mostly Special condition and 62 children in the Mostly Regular condition.

Materials and procedure. The materials, procedure, and design were similar to those in Experiment 1. See Figure 17 for the story and script. In this experiment, however, children saw the balloons at the beginning of the story, and were told that the girl did not know which balloons were in each box. They were asked a comprehension question to confirm they understood this. We also changed the counterfactual question so that it asked whether the girl almost got a special balloon. Children were again tested in either the Mostly Special or the Mostly Regular condition.

Results and Discussion

We first looked at children's happiness ratings. There was a significant main effect of age, *Wald* $X^2(1) = 12.45$, $p < .001$, where younger children rated the girl as happier about receiving a regular balloon than older children did; see Figure 18. There was no effect of distribution, *Wald* $X^2(1) = 2.93$, $p = .087$, and no age-by-distribution interaction, *Wald* $X^2(1) = 0.03$, $p = .857$.

Next, we looked at children's responses to the counterfactual question. There was a significant main effect of age, *Wald* $X^2(1) = 4.55$, $p = .033$, no main effect of distribution, *Wald* $X^2(1) = 2.64$, $p = .104$, and a significant age-by-distribution interaction, *Wald* $X^2(1) = 6.83$, $p = .009$; see Figure 18. We explored each age group separately and found that 6-year-olds were more likely to say the girl almost got a special balloon in the Mostly Regular condition than in the Mostly Special condition, *Wald* $X^2(1) = 3.99$, $p = .046$, while 5-year-olds' responses did not significantly differ between these conditions, *Wald* $X^2(1) = 0.11$, $p = .739$.[1] We also explored each condition separately. In the Mostly Regular condition, with age, children were more likely to say the girl almost got a special balloon, *Wald* $X^2(1) = 9.54$, $p = .002$; in the Mostly Special condition, age did not affect their judgment, *Wald* $X^2(1) = 0.14$, $p = .707$.

In sum, children were insensitive to close counterfactuals when inferring emotions, and did not appropriately differentiate between conditions when answering the counterfactual question. In fact, to the extent that children were sensitive to the difference between conditions for the counterfactual question, it was opposite to expectations. Surprisingly, 6-year-olds were more likely to say that the girl "almost" got a special balloon when this was *less* likely. We are uncertain about what caused this result, but here is one speculation: Although the regular balloon chosen by the girl was next to a special one in both conditions, this might have seemed more remarkable or luckier in the condition where there were just a few special balloons. So in that condition, the low odds of winning a special balloon might have accentuated the closeness of choosing an adjacent balloon.

[1] For 6-year-olds, we had to run this follow-up analysis using a negative binomial distribution because using a binary logistic distribution led to the Hessian matrix being violated. We observed comparable results when we analyzed the data using a Mann-Whitney test ($p = .044$).

Regardless, children's difficulties in this experiment suggest that rather than benefitting children, showing the distribution and locations of balloons in advance hurt their performance. Children may have had difficulty suppressing their knowledge of the balloons' locations when inferring the girl's emotions. If they thought she knew where the balloons were located, they might have thought she chose the regular balloon on purpose, which would affect their emotion inferences. This difficulty is consistent with findings that children, sometimes including 6-year-olds, have difficulty ignoring their privileged knowledge when attributing emotions to others (Bradmetz & Schneider, 1999; de Rosnay et al., 2004; Harris et al., 1989; Ronfard & Harris, 2014).

Alternatively, the finding could be taken to suggest that 6-year-olds' responses in Experiment 1 do not replicate. To address these possibilities, we conducted a further study using the original design in which the participant does not know the initial odds of winning a special balloon. We attempted to make the task easier for children by providing two cues to the closeness of the counterfactual alternative of winning a special balloon. In addition to manipulating probability across the two conditions, we also manipulated spatial proximity. This next experiment also used both close counterfactual questions – half the children were asked whether the girl *could have easily* gotten a special balloon and the other half were asked whether the girl *almost* got a special balloon.

Experiment 3

Children saw a character choose one of ten closed boxes and learn the distribution of the balloons after she had already won a regular one. In this experiment, we also manipulated the position of the special and regular balloons. In the Mostly Special condition, the girl had a high chance and was physically very close to winning a special balloon. In the Mostly Regular condition, the girl had a low chance and was physically far from winning a special balloon.

Methods

Participants. We tested 120 5-6-year-olds (*M* = 5;11; range = 5;0 – 6;11; 55 girls). There were 60 5-year-olds (*M* = 5;4; range = 5;0 – 5;11; 27 girls), and 60 6-year-olds (*M* = 6;6; range = 6;0 – 6;11; 28 girls), with equal numbers of children per age-in-years randomly assigned to each between-subjects condition.

Materials and procedure. The experiment used a 2 x 2 design in which distribution (Mostly Special, Mostly Regular) was manipulated within subjects (order counterbalanced across children) and question-type (easily, almost) was manipulated between subjects. Children each heard two stories. See Figure 17 for a sample of the stories and script. In both stories, a character (girl in story 1, boy in story 2) chose one of ten closed boxes and received a regular balloon. The character was then shown the balloons in the remaining boxes, and children were asked how the character felt about getting the regular balloon. Next, children were either asked whether the character *could have easily* gotten a special balloon, or whether the character *almost* got a special balloon.

The difference between this experiment and the previous experiments is the *position* of each balloon in the distribution. In the previous experiments, in both conditions, the regular balloon that the girl received was flanked by one regular balloon and one special balloon. In this experiment, the regular balloon that the girl received was flanked by two special balloons in the Mostly Special condition, and two regular balloons in the Mostly Regular condition. This could potentially make the happiness and counterfactual questions easier to answer.

Results and Discussion

We first looked at children's happiness ratings. There was a significant main effect of distribution, *Wald* $X^2(1) = 12.23$, $p < .001$, where children rated the character as happier about receiving a regular balloon in the Mostly Regular condition than in the Mostly Special condition; see Figure 18. There was no main effect of age, *Wald* $X^2(1) = 0.78$, $p = .376$, and no age-by-distribution interaction, *Wald* $X^2(1) = 0.78$, $p = .376$.

Next, we looked at children's responses to the counterfactual questions; see Figure 18. There was a significant main effect of distribution, *Wald* $X^2(1) = 14.92$, $p < .001$, as children were overall more likely to say "yes" in the Mostly Special condition than in the Mostly Regular condition (i.e., said the character could have easily gotten or almost got the special balloon), but no main effect of age, *Wald* $X^2(1) = 0.002$, $p = .969$. There was also a significant age-by-distribution interaction, *Wald* $X^2(1) = 7.59$, $p = .006$. This resulted from 6-year-olds being more likely to say "yes" in the Mostly Special condition than in the Mostly Regular condition, *Wald* $X^2(1) = 15.22$, $p < .001$, and 5-year-olds not responding differently between the two conditions, *Wald* $X^2(1) = 1.97$, $p = .161$. Also, in the Mostly Regular condition, with age, children were marginally less likely to say "yes", *Wald* $X^2(1) = 3.70$, $p = .054$; in the Mostly Special condition, age did not affect their judgment, *Wald* $X^2(1) = 0.04$, $p = .834$.

Further, there was a significant main effect of question-type, *Wald* $X^2(1) = 13.82$, $p < .001$, where children were overall more likely to say "yes" when asked whether the character almost got a special balloon. Although we collapsed across question-type when showing these results in Figure 18, we show the results split by question-type in Appendix A. There was also a significant age-by-question-type interaction, *Wald* $X^2(1) = 9.92$, $p = .002$. This resulted because 6-year-olds were more likely to say "yes" when asked whether the character almost got a special balloon than when asked whether the character could have easily gotten a special balloon, *Wald* $X^2(1) = 19.89$, $p < .001$, whereas 5-year-olds' responses did not differ between the two question-types, *Wald* $X^2(1) = 0.81$, $p = .369$. Also, when asked about whether the character could have easily gotten a special balloon, with age, children were less likely to say "yes", *Wald* $X^2(1) = 5.14$, $p = .023$; when asked about whether the character almost got a special balloon, age did not affect their responses, *Wald* $X^2(1) = 2.57$, $p = .109$. Because these analyses do not involve distribution, our main factor of interest, we do not discuss them further.

Finally, we looked at whether children's responses to the counterfactual questions corresponded with their happiness ratings. If children reason about close counterfactuals, they should judge that the girl almost (or could have easily) received a special balloon in the Mostly Special condition, but not in the Mostly Regular condition. Overall, 25 children showed this pattern, and 95 did not. We used a GEE model (ordinal logistic) to test whether this pattern of responding predicted whether children gave more positive emotion ratings in the Mostly Regular condition than in the Mostly Special condition. The model included distribution (Mostly Special, Mostly Regular) and whether children showed this pattern of close-counterfactual reasoning as factors. There was a significant main effect of distribution, *Wald* $X^2(1) = 4.89$, $p = .027$, and no main effect of pattern-type, *Wald* $X^2(1) = 3.45$, $p = .063$. There was also no distribution-by-pattern-type interaction, *Wald* $X^2(1) = 0.85$, $p = .357$, suggesting that children's happiness ratings did not differ based on how they responded to the counterfactual questions.

In sum, both 5- and 6-year-olds inferred the girl would feel sadder about winning a regular balloon in the condition where she was closer to winning a better one. Further, 6-year-olds also responded correctly to the close counterfactual questions. However, children's success in inferring emotions did not depend on their success in close counterfactual reasoning, suggesting that they may not use close counterfactuals to infer emotions. This raises the concern that even adults' close counterfactual reasoning might not correspond with their emotion predictions. We assessed this possibility in our final experiment.

Experiment 4

We tested adults using a similar design to Experiment 3. Our goal was to have a reference point with which to compare children's responses and to address the concern that previous studies on adults confounded close counterfactuals with prior beliefs (e.g., Mellers et al., 1997; van Dijk & van der Pligt, 1997).

Methods

Preregistration. This experiment was preregistered at https://aspredicted.org/blind.php?x=829yz4. The preregistration covered the number of participants, design, and analysis plan.

Participants. The experiment was successfully completed by 227 adults (mean age = 35 years; 74 females) recruited using Amazon Mechanical Turk. Participants received $0.75 USD for their participation. We originally had 302 participants, but as per our preregistration, we excluded those who failed at least one of our two comprehension check questions or neglected to answer any of the test questions ($n = 75$).

Materials and procedure. The experiment used the same 2 x 2 design as Experiment 3: distribution (Mostly Special, Mostly Regular) was manipulated within subjects, and question-type (easily, almost) was manipulated between subjects. The stories were similar to those in Experiment 3, except the language of the script was changed to be more suitable for adults. The first page showed a girl in front of two balloons with the script, "Here's a girl. She can win a balloon. She might win a regular balloon (left) or she might win a special balloon (right)". The second page showed the girl in front of ten closed boxes with the script, “The girl gets to choose a box and she’ll win whichever balloon is in the box she chooses”. The third page showed the girl receiving a regular balloon with the script, “The girl chooses this box and wins this balloon. Let’s show the girl what balloons are in the other boxes”. The fourth page showed the remaining balloons. The test question (“When the girl sees this, how does she feel about the balloon she won?”) appeared beneath the image, along with the same happy face scale used with children. Participants responded by clicking on one of the faces. Participants then saw a story about a boy with the other distribution. The counterfactual question for both stories was asked after participants had already seen and made happiness judgments for both distributions (e.g., “Here’s the girl and all of the balloons again. Could the girl have easily won a special balloon?” or “Here’s the girl and all of the balloons again. Did the girl almost win a special balloon?”). We left these counterfactual questions for the end of the procedure (i.e., rather than asking a counterfactual question after each story) because we were concerned that asking the counterfactual question after the first story might affect how adults approached the emotion judgment in the second story.

Results and Discussion

We first looked at adults’ happiness ratings. There was a significant main effect of distribution, *Wald* $X^2(1) = 72.84$, $p < .001$, where they rated the character as happier about receiving a regular balloon in the Mostly Regular condition ($M = 0.28$; $SD = 1.38$) than in the Mostly Special condition ($M = -1.12$; $SD = 1.89$).

Next, we looked at adults’ responses to the counterfactual questions. There was a significant main effect of distribution, *Wald* $X^2(1) = 177.77$, $p < .001$, where they were more likely to respond “yes” in the Mostly Special condition (i.e., said the character could have easily gotten or almost got the special balloon). There was no main effect of question-type, *Wald* $X^2(1) = 3.16$, $p = .076$, and no distribution-by-question-type interaction, *Wald* $X^2(1) = 1.09$, $p = .296$.

Finally, we looked at how adults' responses to the counterfactual questions corresponded with their happiness ratings. Responses were coded in the same manner as in Experiment 3. Overall, 174 participants showed the expected pattern ("yes" in the Mostly Special condition and "no" in the Mostly Regular condition), and 53 did not. We conducted a GEE model (ordinal logistic) with distribution and pattern-type entered into the model as factors. There was a significant main effect of distribution, *Wald* $X^2(1) = 30.73$, $p < .001$, a significant main effect of pattern-type, *Wald* $X^2(1) = 14.13$, $p < .001$, and a distribution-by-pattern-type interaction, *Wald* $X^2(1) = 13.67$, $p < .001$. Participants who responded both "yes" in the Mostly Special condition and "no" in the Mostly Regular condition rated the character as happier about receiving a regular balloon in the Mostly Regular condition than in the Mostly Special condition, *Wald* $X^2(1) = 82.12$, $p < .001$. Participants who responded to the counterfactual questions in any other way showed no differences in happiness ratings between the two conditions, *Wald* $X^2(1) = 1.43$, $p = .231$.[2] Unlike children, adults' emotion ratings corresponded with their close counterfactual responses, suggesting that they consider close counterfactual alternatives when inferring emotions.

General Discussion

We investigated young children's ability to consider close counterfactual alternatives when inferring other people's emotions. We first examined counterfactual closeness in connection with probability and found that 6-year-olds (but not younger children) recognized

[2] In our pre-registration, we reported that we would run the analysis for the happiness ratings using a paired-samples t-test (we report GEEs in the main text because they are more appropriate for this ordinal data). The results are essentially the same. See Appendix B for the pre-registered analyses.

that a person would be less happy about receiving a regular prize when the prior probability of receiving a special prize was high rather than low. However, children did not show this understanding when they knew the odds of winning a special prize from the outset. We next examined counterfactual closeness in connection with probability and spatial proximity combined. Now 5- and 6-year-olds, and adults, recognized that the person would be less happy about winning a regular prize when the possibility of receiving a special prize was a close counterfactual alternative.

In each experiment, participants also responded to counterfactual questions about whether the alternative outcome of winning a special prize was close or far. In the first two experiments, children did not appropriately respond to these questions. They were no more likely to say the girl could easily have won a special balloon (or that she almost won one) when the prior probability of this alternative outcome was high rather than low. In the final experiments (where counterfactual closeness depended on both prior probabilities and spatial proximity), 6-year-olds (but not 5-year-olds) and adults responded appropriately. But whereas adults' responses to the close counterfactual questions predicted their emotion inferences, 6-year-olds' responses did not.

Together these findings provide conflicting evidence about whether children consider close counterfactuals when inferring emotions. The emotion ratings suggest that by age 5 or 6, children do consider close counterfactuals in these inferences – their emotion judgments were broadly in line with those of adults in classic studies connecting close counterfactuals and emotions (e.g., Johnson, 1986; Mellers et al., 1997; van Dijk & van der Pligt, 1997). At the same time, children's largely poor performance in response to the counterfactual questions provides reason to doubt this conclusion.

One interpretation of the findings, then, is that children did *not* use close counterfactuals to infer emotions. To be viable, though, this account needs to specify how else children could have made these inferences. We do not see a plausible alternative. For example, it might seem that children could have inferred the girl's emotions by considering her prior beliefs and expectations (e.g., Asaba et al., 2019; Lara et al., 2019; Wellman & Bartsch, 1988; also see Wu & Schulz, 2018). But this is unlikely because in Experiments 1 and 3, children had no information about the girl's prior beliefs (i.e., the distribution of balloons was revealed *after* the girl had already won a regular one).[3] It might also seem that children could have inferred emotions by using the distribution of balloons as a standard of comparison (Cadotte et al., 1987; also see Doan et al., 2020). For example, when most balloons are special, winning a special balloon may be viewed as the standard outcome. Receiving a regular balloon is disappointing because it falls short of this standard. But this explanation cannot explain why children's judgments should be sensitive to the spatial arrangements of the regular and special balloons, and it appears that children were sensitive to this factor (i.e., 5-year-olds performed better when counterfactual closeness depended on spatial proximity and probability together, compared with when it depended on probability alone).

We instead favour an account holding that children *do* use close counterfactuals to infer emotions, but that our participants had difficulty understanding and responding to the counterfactual questions. Several factors could have contributed to this difficulty. For example, the "easily could have" question is somewhat figurative, and children might have had difficulty understanding it; no previous research had used this question with children. Previous work has

[3] Further, the findings from Experiment 2 suggest that information about prior beliefs interfered with performance. When children saw the distribution and locations of the balloons at the beginning, they had difficulty ignoring their privileged knowledge and did not rate the girl's emotions differently between conditions (for related findings, see Bradmetz & Schneider, 1999; de Rosnay et al., 2004; Harris et al., 1989; Ronfard & Harris, 2014).

found that 5-year-olds do understand counterfactual questions about what "almost happened" (Beck & Guthrie, 2011). But children might need to see contrasting events to successfully answer this kind of question. In Beck and Guthrie's experiments, after seeing two toy horses galloping along a table, 5-year-olds said the one who stopped right at the edge almost fell, as compared to either one that fell or one that did not come close to falling (Beck & Guthrie, 2011). Perhaps children needed to compare the horse who stopped at the edge with one who fell or did not come close to falling in order to determine which one almost fell. Finally, our participants may have struggled with the way our questions were ordered. Perhaps they would have found the counterfactual questions easier if these had been asked immediately after the scenario, and not after questions about the girl's emotion.

Our primary interest in conducting this research was examining children's use of close counterfactuals when inferring emotions. As discussed, though, counterfactual closeness depends on many factors, including probability, spatial proximity, and numerical proximity. Our first experiments manipulated closeness via probability alone, and our later experiments manipulated it alongside spatial proximity. Differences between children's responses in Experiments 1 and 3 suggest that children could be more sensitive to spatial proximity than probability as a cue to counterfactual closeness, and that spatial proximity may precede probability as a cue to counterfactual closeness. However, these conclusions are tentative at best. Adequately investigating it will require a single experiment that separately manipulates both factors (i.e., rather than making informal comparisons across experiments). It is also possible that future research might reveal deeper differences between these determinants of counterfactual closeness. Indeed, it is even possible that probability-based and spatially-based closeness could turn out to be different kinds of counterfactual closeness. Returning to our counterfactual questions, while it might be appropriate to say that an outcome with a high prior probability could have easily happened, it might be less appropriate to say it almost or nearly happened.

In summary, we found that by age 5 or 6, children recognize that another person will be sadder about an outcome if a better outcome could easily have occurred, rather than if a better outcome could not have easily occurred. These emotion judgments contrast with children's judgments in earlier work, which suggested that before age 7, children do not consider counterfactual possibilities when inferring emotions (e.g., Amsel & Smalley, 2000; Beck & Crilly, 2009; Guttentag & Ferrell, 2004). In our experiments, children largely had difficulty assessing whether the better outcome was a close counterfactual, even when they were successful in these emotion inferences. Although further work is needed, we suspect that difficulty understanding these counterfactual questions masked children's reasoning about close counterfactuals.

Chapter Five: General Discussion

Summary of Findings

In 13 experiments, I examined the development of children's (and adults') ability to consider probability when making emotion inferences. Chapter Two investigated children's use of probabilistic information when inferring other people's surprise. Through four experiments, I found development in children's ability to use probability to infer surprise. At age 5 and under, children do not use probability to infer surprise. At age 6, children have a limited ability to use probability to infer surprise, such that they require prompts regarding probability (but not beliefs) in order to consistently use it to infer surprise. It is only at age 7 that children have a robust understanding of the connection between probability and surprise. These findings are the first to show that children consider probability when attributing surprise to others. They also demonstrate that children's understanding of surprise is not just belief-based, but also probability-based.

Chapter Three investigated children's use of probabilistic information when inferring the quality of outcomes and other people's happiness. Through four experiments, I found a lag between children's ability to use probability to infer outcome quality and their ability to use probability to infer happiness. By age 4, children understand that the quality of an outcome is influenced by whether a better outcome was initially more likely. However, it is not until age 5 that children understand that people's happiness with an outcome is also dependent on the likelihood of a better outcome. A fifth experiment showed that adults use probability to assess both quality and happiness. These findings are the first to show that children consider probability when inferring others' happiness.

Chapter Four investigated children's use of close counterfactual alternatives (manipulated through probability) when inferring other people's happiness. Through three experiments, I found development in children's ability to consider the closeness of alternative outcomes to infer happiness. At age 6, children recognize that people's happiness with an outcome is dependent on whether a better outcome was initially more likely. Children recognize this at age 5 if the better outcome was both initially more likely and physically close to the person. Further, at age 6, children acknowledge that winning a better prize was a likely counterfactual alternative when the better prize was more likely and physically close, however their close counterfactual reasoning do not predict their emotion inferences. A fourth experiment showed that adults' close counterfactual reasoning predicts their emotion inferences. These findings are the first to show how close counterfactual alternatives, manipulated through probability, influences children's happiness inferences.

The findings from these three chapters provide the first evidence that children use probability to infer other people's emotions. They also provide a developmental view of when children integrate their understanding of probability with their understanding of emotions. In the remainder of this chapter, I discuss how the findings from my research (do not) fit into the current accounts of how children infer emotions, and theorize about how children might be using probability to infer emotions. I also discuss the inconsistencies of the statistical analyses between my chapters, and the limitations of my work and possible future directions.

Beyond Current Accounts of Children's Emotion Understanding

The findings of my dissertation contrast with existing accounts of how children infer others' emotions. One major account proposes that children infer emotions by learning and memorizing scripts (e.g., Widen & Russell, 2010, 2011; also see Fehr & Russell, 1984). My findings are not easily handled by script theory. Scripts are composed of specific and concrete events (e.g., Abelson, 1981; Fehr & Russell, 1984; Schank & Abelson, 1975), meaning that scripts do not include general and abstract notions (Gopnik & Meltzoff, 1997). However, probabilities are not specific events, and the notion of probability is abstract rather than concrete. For example, although it might be easy to depict or visualize specific high probability (or low probability) events, there is no way to do this for the abstract notion of being *highly probable*. So probability likely does not enter into scripts.

One might posit, though, that script theory could be adapted or extended to include abstract and general concepts like probability. For example, one could propose that children's script for surprise could be something like, "when unlikely events occur, people are surprised". However, this is better described as a folk or naïve theory of surprise, given that these knowledge structures are not limited to specific events, and instead typically reference abstract and non-obvious concepts (e.g., Gelman & Noles, 2011). So suggesting that scripts can include abstract concepts like probabilities would collapse the distinction between scripts and other kinds of knowledge structures. This revision to the script theory of emotions would also render the theory unfalsifiable and empty. If scripts can include *any* information useful for inferring emotions, then no matter what information children use, the script theory is supported.

Another major account suggests that children infer emotions by drawing on their "theory of mind" and considering others' mental states, like their desires and beliefs (e.g., Harris, 2008; Rieffe et al., 2005; Wellman & Banerjee, 1991). My findings also do not fit directly with this account. When people reason about probability, they do not *need* to consider mental states. For example, you can conclude that someone who is at a gumball machine with many purple gumballs and few red gumballs is unlikely to get a red one without considering their beliefs or desires about what they would receive before the outcome occurs. Further, probabilistic information is often enough to allow us to infer emotions. The knowledge that the person is unlikely to receive a red gumball is enough to allow us to infer that they will be surprised if they get one.

However, it could *seem* that this inference requires us to also consider their beliefs or knowledge. For example, it might seem that we inferred their surprise by reasoning as follows: They know they are unlikely to get a red gumball, so they believe they will not get one; therefore, they are surprised when they do get one. We *could* infer surprise in this way. However, my findings from Chapter Two suggests that children are not inferring surprise in this way. Experiments 3 and 4, in which children were provided with a probability, belief, or control prompt, show that children do not need to consider other people's beliefs to infer their surprise. In fact, I found that prompting children to consider beliefs did not improve their surprise inferences, but prompting them to consider probability did. As such, in cases like this, probability information is sufficient to infer surprise.

One might wonder whether children consider mental states when inferring happiness. In Chapters Three and Four, children must of course consider the girl's desires when inferring her happiness. But children might also have considered her beliefs. For example, in Chapter Three, when the gumball machine contained mostly yummy gumballs, children might have reasoned that the girl was somewhat disappointed to get equal numbers of yummy and yucky gumballs because she had *believed* she would get mostly yummy ones. On this account, children may have used probabilistic reasoning to infer the girl's belief about what she would receive, but then used this belief attribution (instead of the probability information) to infer her emotion. That is, because there are many yummy gumballs, the girl believed she would get mostly yummy ones, and because she believed she would get mostly yummy ones, she is disappointed when she does not. As such, children might have made their happiness judgments based on the girl's beliefs. Chapter Four allowed me to examine this possibility more closely. Children saw a girl choose one of ten closed boxes and receive a regular balloon. In one condition, the girl later learned that most of the other boxes contained regular balloons, while in another condition, she learned that most of them held special balloons. If children's inferences of happiness are belief-based, then their happiness rating should not differ across these two conditions – when the girl chose a box, she was unaware of the distribution of regular and special balloons, and so her prior beliefs about what she would receive could not differ by condition. However, children's happiness judgments *did* vary across the two conditions in Experiments 1 and 3. They judged that the girl was happy with the regular balloon she received when most of the other balloons were regular, but they judged she felt sad about it when most of the other balloons were special. This pattern shows that children consider probability when inferring others' happiness, as they judged that the girl was happier when she discovered (after-the-fact) that she had had a low chance of getting a special balloon. So, although people may sometimes consider people's belief-based mental states alongside probability when inferring emotions, it is not always necessary – considering probability by itself is often enough.

Together, my findings suggest that the dominant accounts of how children infer emotions do not suffice in accounting for the different ways in which children can infer emotions. Specifically, they cannot accommodate the findings that children use probability to infer other people's emotions. Children's use of probability shows that they can infer emotions by considering abstract concepts from outside the domain of theory of mind. This conclusion is also supported by research showing that children infer emotions by drawing on their understanding of ownership (Pesowski & Friedman, 2015).

This raises the question of what a theory of emotion should entail. Research in social cognition has shown that computational Bayesian models can be applied to people's inferences about the social world (e.g., Baker et al., 2017; Jara-Ettinger et al., 2016; Jern & Kemp, 2015). Recent work has applied these computational cognitive modeling approaches to specifically model people's understanding of emotions (e.g., Ong et al., 2015; Wu et al., 2018; see Ong et al., 2019 for a review). An intuitive theory of emotion has been proposed to unify various lay reasoning about emotions into one cohesive framework. Much like how we use multiple senses (e.g., sight, smell, touch) to understand our physical world, we can draw on a variety of cues from the environment and from people when inferring emotions. For example, we can integrate information about the outcome itself, the person's mental states, and the person's facial expressions to infer their emotion. An intuitive model of emotion allows for this integration and does not assume *a priori* that one cue would be weighted more heavily than another – the model assumes that people are sensitive to the reliability of each cue in certain contexts and allows for differing weights of each cue depending on the context (Ong et al., 2019). I propose that information about probability and statistics can also be integrated into this model such that we can use information about probability in addition to the other cues and factors to infer emotions. As I have shown in the previous three chapters, information about probability can be highly relevant and is used in both children's and adults' inferences of emotions. Further, there are instances where probability information may be more relevant than information about mental states. That said, when multiple cues are available, I suspect that how we weigh probability information in comparison to other cues is highly dependent on past experiences and the situational context. Overall, the intuitive theory framework for studying how people reason about emotions may be the much needed bridge that connects our lay reasoning about emotions together.

How do Children use Probability to Infer Emotions?

Perhaps the most important question raised by this work is *how* children use probability to infer emotions. One explanation is that these inferences depend on theory-like generalizations connecting probability with different emotions. This account may be plausible for children's inferences of surprise, as they could depend on a simple and intuitive generalization like *people are surprised when improbable events happen.* However, it is harder to see how theory-like generalizations could support children's inferences of happiness, as this would require children to use less intuitive generalizations like *people's happiness about positive outcomes increases as the likelihood of the outcomes decreases.* In fact, this generalization is unlikely to explain children's responses in my research, as they involved outcomes that were not clearly positive or negative (e.g., receiving two yummy gumballs and two yucky ones).

However, children's probability-based assessments of happiness could depend on generalizations nonetheless. Children might reason according to the intuitive generalization that *more positive outcomes increase people's happiness (and more negative outcomes decrease it)*. This generalization does not involve probability. However, it does require children to think about the extent to which outcomes are positive or negative, and children can accomplish this through probabilistic reasoning. Children might have engaged in this process in Chapters Three and Four. For example, in Chapter Three, children could assess that receiving two yummy and two yucky gumballs is bad (or unlucky) if they came from a gumball machine that contained mostly yummy gumballs. Then, because the outcome is considered bad, children would infer that the girl who received this outcome is sad. Similarly, in Chapter Four, after the distribution of balloons were revealed, children could assess that getting a regular balloon is bad when most of the other boxes contained special balloons, thus inferring the girl is sad. As such, inferences of happiness or sadness might require two steps: (1) use probability to assess the quality of the outcome, and (2) use the quality of the outcome to infer happiness. My findings from Chapter Three demonstrate that children can use probability to assess the quality of an outcome by age 4, but do not use it to infer happiness until age 5. It is possible that when inferring happiness, 4-year-olds do not spontaneously do the first step, thus they are unable to successfully infer happiness and sadness.

If children infer emotions in the way that I have proposed, it would suggest that there are different paths by which probability enters into children's inferences of others' emotions. To infer surprise, children may need to represent generalizations like *improbable events are surprising*, which directly link the conceptual domains of emotion and probability. Younger children may actually struggle with this direct link. However, to infer happiness from probability, children might not need to directly link these domains. Thus, there might be a direct path for inferences of surprise (probability information → surprise), and an indirect path for inferences of happiness (probability information → quality → happiness). And even though inferring happiness from probability might take more steps than inferring surprise from probability, these steps are arguably simpler, allowing children to infer happiness at a younger age.

Consistency of Analyses

There is inconsistency in the statistical analyses that I conducted between each chapter of my dissertation (i.e., GEEs in Chapters Two and Four and ANOVAs in Chapter Three, age in years in Chapters Two and Three and age in months in Chapter Four). My dissertation consists of three published or under review manuscripts and the analyses in each paper are a reflection of the statistical tests that I was aware of at the time the manuscripts were submitted. However, throughout my PhD, the conventions for analyzing data have begun to shift and I have learned new analyses that may be more appropriate for the type of data that is in my dissertation. Thus, in Appendix C, I have re-analyzed my data for Chapters Two and Three so that my statistical analyses are consistent across the three chapters and are in accord with the current standards. Specifically, I conducted GEEs for all analyses and treated age as a continuous co-variate (age in months). I conducted GEEs because unlike ANOVAs, GEEs can deal with categorical dependent variables, like the forced-choice binary and ordinal responses that I have. Further, analyzing the data with children's age in months allows me to examine the developmental changes as a continuum as opposed to observing differences between each age in years. This is a more sensitive test and better represents the samples of children in my experiments, given that all experiments included continuous age ranges, rather than discontinuous years (i.e., I generally tested 4 – 6 year olds and not something like a group of 4-year-olds, a group of 7-year-olds and a group of 10-year-olds). The findings from these new analyses are fairly consistent with the original analyses. Thus, the interpretation of my findings in the dissertation do not change. The consistency of the results between the two different analyses also increase my confidence in the robustness of the effects.

Limitations

One limitation of my dissertation is that I have yet to nail down the exact mechanism in which children are using probability to infer emotions. I have proposed that there might be different pathways for inferences of surprise and happiness, but I have not tested these possibilities directly. In the future, I can test my proposed pathway for happiness inferences by conducting prompting studies. That is, if children first consider probability to infer the quality of an outcome, and then use the quality of the outcome to infer happiness, then prompting children to consider outcome quality before they make their emotion judgment should improve their happiness inferences. Relatedly, if children assess the extent to which an outcome is positive or negative by considering the likelihood of better or worse outcomes, then this suggests that they might consider counterfactual possibilities or use probability as a standard of comparison. However, my findings from Chapter Four provides mixed evidence and cannot speak to the exact mechanism. Further, it is possible that different mechanisms are at play in different situations. As such, much more work is required to determine how children are making these inferences.

Another limitation is that in all experiments, I tested participants from a Western, educated, industrialized, rich, and democratic (WEIRD; Henrich et al., 2010) society. As such, we must not assume that these results would generalize to non-WEIRD populations. Also, most of my experiments used similar stimuli involving gumball machines. In choosing gumball machines, I anticipated that children would be familiar with their probabilistic nature. However, it is uncertain whether my findings would replicate using other kinds of probabilistic stimuli (e.g., spinners, dice).

Future Directions

The current findings provide the first evidence that children use probability to infer other people's emotions. While these findings broaden our understanding of children's conceptions of emotion, much future work is needed in order to capture the whole picture. In this section, I discuss some possible future directions.

To fully understand the connection between probability and emotions, children should be able to reason flexibly back and forth between the two concepts. That is, not only should children be able to use probability to infer emotions, they should also be able to make "backward" inferences about the probabilistic context based on emotions. Follow-up work could examine this question directly. For example, one could show children a girl holding a red gumball, standing between two gumball machines. One machine has many red gumballs and few purple ones, and the other machine has many purple gumballs and few red ones. When told that the girl is very surprised that she got a red gumball, will children infer that the girl got her gumball from the machine with fewer red ones? We could also ask a similar question regarding disappointment. One could show children a girl holding two red yummy and two black yucky gumballs, standing between two gumball machines. One machine has many yummy gumballs and few yucky ones, and the other machine has many yucky gumballs and few yummy ones. When told that the girl is very disappointed about the outcome, will children infer that she got her gumballs from the machine with mostly yummy gumballs? Examining these questions will allow us to better understand how children's understandings of probability and emotion influence each other.

Furthermore, to fully understand the connection between probability and emotions, we must examine various emotions. In my dissertation, I focused only on children's inferences of surprise and happiness/sadness. However, probability affects other emotions. For example, in contests involving skill, we may feel *prouder* of winning when our odds of winning are very low, and we might likewise feel more *ashamed* of losing if our success seemed all but guaranteed. Probability also affects fear. For example, people worry more about a dangerous or negative outcome when they believe it is more probable (e.g., Baron et al., 2000; Peters et al., 2006). The relation between fear and probability is complicated, though, as people's assessments of the likelihoods of some events are wildly inaccurate and fearing an event can make people overlook the fact that the likelihood of it occurring is low (e.g., Sunstein 2002, 2003). Nevertheless, the effects of probability on emotions like pride, shame, and fear raises the possibility that children might infer these emotions by considering probability.

Future research should also aim at uncovering strategies that may improve children's emotion inferences. The findings from Chapter Two (Experiments 3 and 4) suggest that prompting children to consider probability is one way to improve their surprise inferences. This may also be true for other emotion inferences. Before children are asked to make their emotion inferences regarding a particular outcome, they could be asked to consider whether the chance of that outcome occurring is high or low. If these prompting strategies improve children's emotion inferences, it would suggest that they just need reminders to consider probability. Follow-up work could investigate whether training paradigms or interventions aimed at improving children's probabilistic reasoning would also improve their emotion inferences. That is, if children are taught to consider probability in their daily lives, they might come to spontaneously use it when inferring emotions.

Conclusions

Overall, my dissertation provides novel insights into children's emotion understanding. My findings show development in children's ability to integrate probabilistic information into their various emotion inferences. At age 4, children use probability to infer the quality of outcomes; at age 5, they use probability to infer other people's happiness; at age 6, they use probability to infer other people's surprise when prompted; and at age 7, they spontaneously use probability to infer surprise. These findings broaden our understanding of children's conceptions of emotions by revealing another way in which children can infer other people's emotions. These findings are also informative about children's probabilistic reasoning and how they use probability to make social inferences.

References

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist, 36*(7), 715-729. doi: 10.1037/0003-066X.36.7.715

Amsel, E., & Smalley, J. D. (2000). Beyond really and truly: Children’s counterfactual thinking about pretend and possible worlds. In K. Riggs & P. Mitchell (Eds.), *Children’s reasoning and the mind* (pp.121-147). Hove, UK: Psychology Press.

Asaba, M., Ong, D. C., & Gweon, H. (2019). Integrating expectations and outcomes: Preschoolers’ developing ability to reason about others’ emotions. *Developmental Psychology, 55*(8), 1680-1693. doi: 10.1037/dev0000749

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321-324. doi: 10.1111/1467-9280.00063

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 1-10. doi: 10.1038/s41562-017-0064

Barden, R. C., Zelko, F. A., Duncan, S. W., & Masters, J. C. (1980). Children's consensual knowledge about the experiential determinants of emotion. *Journal of Personality and Social Psychology*, *39*(5), 968-976. doi: 10.1037/0022-3514.39.5.968

Baron, J., Hershey, J. C., & Kunreuther, H. (2000). Determinants of priority for risk reduction: The role of worry. *Risk Analysis, 20*(4), 413-428. doi: 10.1111/0272-4332.204041

Bartsch, K., & Estes, D. (1997). Children's and adults' everyday talk about surprise. *British Journal of Developmental Psychology*, *15*(4), 461-475. doi: 10.1111/j.2044-835X.1997.tb00741.x

Beck, S. R., & Crilly, M. (2009). Is understanding regret dependent on developments in counterfactual thinking?. *British Journal of Developmental Psychology*, *27*(2), 505-510. doi: 10.1348/026151008X401697

Beck, S. R., & Guthrie, C. (2011). Almost thinking counterfactually: Children’s understanding of close counterfactuals. *Child Development*, *82*(4), 1189-1198. doi: 10.1111/j.1467-8624.2011.01590.x

Beck, S. R., & Riggs, K. J. (2014). Developing thoughts about what might have been. *Child Development Perspectives, 8*(3), 175-179. doi: 10.1111/cdep.12082

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2010). The effect of causal chain length on counterfactual conditional reasoning. *British Journal of Developmental Psychology*, *28*(3), 505-521. doi: 10.1348/026151009X450836

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, *77*(2), 413-426. doi: 10.1111/j.1467-8624.2006.00879.x

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations Research*, 33(1), 1-27. doi: 10.1287/opre.33.1.1

Blake, P. R., & McAuliffe, K. (2011). “I had so much it didn’t seem fair”: Eight-year-olds reject two forms of inequity. *Cognition*, *120*(2), 215-224. doi: 10.1016/j.cognition.2011.04.006

Borke, H. (1971). Interpersonal perception of young children: Egocentrism or empathy?. *Developmental Psychology*, *5*(2), 263-269. doi: 10.1037/h0031267

Boyce, C. J., Brown, G. D., & Moore, S. C. (2010). Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological Science*, *21*(4), 471-475. doi: 10.1177/0956797610362671

Bradmetz, J., & Schneider, R. (1999). Is Little Red Riding Hood afraid of her grandmother? Cognitive vs. emotional response to a false belief. *British Journal of Developmental Psychology*, *17*(4), 501-514. doi: 10.1348/026151099165438

Bullock, M., & Russell, J. A. (1985). Further evidence on preschoolers' interpretation of facial expressions. *International Journal of Behavioral Development*, *8*(1), 15-38. doi: 10.1177/016502548500800103

Cadotte, E. R., Woodruff, R. B., & Jenkins, R. L. (1987). Expectations and norms in models of consumer satisfaction. *Journal of Marketing Research*, 305-314. doi: 10.2307/3151641

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595-616. doi: 10.1207/s15326942dn2802_3

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*(4), 1032-1053. doi: 10.1111/1467-8624.00333

Carlson, S. M., & Wang, T. S. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development*, *22*(4), 489-510. doi: 10.1016/j.cogdev.2007.08.002

Cassidy, J., Parke, R. D., Butkovsky, L., & Braungart, J. M. (1992). Family-peer connections: The roles of emotional expressiveness within the family and children's understanding of emotions. *Child Development*, *63*(3), 603-618. doi: 10.1111/j.1467-8624.1992.tb01649.x

De Rosnay, M., Pons, F., Harris, P. L., & Morrell, J. M. (2004). A lag between understanding false belief and emotion attribution in young children: Relationships with linguistic ability and mothers' mental-state language. *British Journal of Developmental Psychology*, *22*(2), 197-218. doi: 10.1348/026151004323044573

Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers: Contextual validation. *Child Development*, 194-201. doi: 10.2307/1130651

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in children's causal inferences: The sampling hypothesis. *Cognition*, *126*(2), 285-300. doi: 10.1016/j.cognition.2012.10.010

Denison, S., Konopczynski, K., Garcia, V., & Xu, F. (2006). Probabilistic reasoning in preschoolers: Random sampling and base rate. In R. Sun & N. Miyake (Eds.), Proceedings of the 28th annual conference of the Cognitive Science Society (pp. 1216–1221). Florence, KY: Psychology Press.

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. doi: 10.1037/a0028278

Denison, S., & Xu, F. (2010). Twelve-to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, *13*(5), 798-803. doi: 10.1111/j.1467-7687.2009.00943.x

Diesendruck, G., Salzer, S., Kushnir, T., & Xu, F. (2015). When choices are not personal: The effect of statistical and social cues on children's inferences about the scope of preferences. *Journal of Cognition and Development*, *16*(2), 370-380. doi: 10.1080/15248372.2013.848870

Doan, T., Friedman, O., & Denison, S. (2018). Beyond belief: The probability-based notion of surprise in children. *Emotion*. doi: 10.1037/emo0000394

Doan, T., Friedman, O., & Denison, S. (2020). Young children use probability to infer happiness and the quality of outcomes. *Psychological Science*, *31*(2), 149-159. doi: 10.1177/0956797619895282

Donaldson, S. K., & Westerman, M. A. (1986). Development of children's understanding of ambivalence and causal theories of emotions. *Developmental Psychology*, *22*(5), 655-662. doi: 10.1037/0012-1649.22.5.655

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, *113*(3), 464. doi: 10.1037/0096-3445.113.3.464

Gelman, S. A., & Noles, N. S. (2011). Domains and naïve theories. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 490-502. doi: 10.1002/wcs.124

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science*, *6*(5), 514-523. doi: 10.1111/1467-7687.00309

Girotto, V., Fontanari, L., Gonzalez, M., Vallortigara, G., & Blaye, A. (2016). Young children do not succeed in choice tasks that imply evaluating chances. *Cognition*, *152*, 32-39. doi: 10.1016/j.cognition.2016.03.010

Gopnik, A. & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Sciences*, *8*(8), 371-377. doi: 10.1016/j.tics.2004.06.005

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620-629. doi: 10.1037/0012-1649.37.5.620

Gove, F. L., & Keating, D. P. (1979). Empathic role-taking precursors. *Developmental Psychology*, *15*(6), 594-600. doi: 10.1037/0012-1649.15.6.594

Guerini, R., FitzGibbon, L., & Coricelli, G. (in press). The role of agency in regret and relief in 3-to 10-year-old children. *Journal of Economic Behavior & Organization*. doi: 10.1016/j.jebo.2020.03.029

Guttentag, R., & Ferrell, J. (2004). Reality compared with its alternatives: age differences in judgments of regret and relief. *Developmental Psychology*, *40*(5), 764. doi: 10.1037/0012-1649.40.5.764

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, *107*(20), 9066-9071. doi: 10.1073/pnas.1003095107

Hadwin, J., & Perner, J. (1991). Pleased and surprised: Children's cognitive theory of emotion. *British Journal of Developmental Psychology*, *9*(2), 215-234. doi: 10.1111/j.2044-835X.1991.tb00872.x

Harris, P.L. (2008). Children’s understanding of emotion. In Lewis, M., Haviland-Jones, J. M., & Barrett, L. F. (Eds.), *Handbook of Emotions* (320-331). Guilford Press.

Harris, P. L., de Rosnay, M., & Pons, F. (2016). Understanding emotion. In Barrett, L. F. , Lewis, M., & Haviland-Jones, J. M.(Eds.), *Handbook of emotions* (293-306). Guilford Press.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, *61*(3), 233-259. doi: 10.1016/S0010-0277(96)00715-9

Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition & Emotion*, *3*(4), 379-400. doi: 10.1080/02699938908412713

Harris, P. L., Olthof, T., Terwogt, M. M., & Hardman, C. E. (1987). Children's knowledge of the situations that provoke emotion. *International Journal of Behavioral Development*, *10*(3), 319-343. doi: 10.1177/016502548701000304

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61-83. doi: 10.1017/S0140525X0999152X

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: longitudinal associations with mental-state talk between young friends. *Developmental Psychology*, *34*(5), 1026-1037. doi: 10.1037/0012-1649.34.5.1026

Hughes Jr, R., Tingle, B. A., & Sawin, D. B. (1981). Development of empathic understanding in children. *Child Development*, 122-128. doi: 10.2307/1129221

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589-604. doi: 10.1016/j.tics.2016.05.011

Jern, A., & Kemp, C. (2015). A decision network account of reasoning about other people's choices. *Cognition*, *142*, 12-38. doi: 10.1016/j.cognition.2015.05.006

Johnson, J. T. (1986). The knowledge of what might have been: Affective and attributional consequences of near outcomes. *Personality and Social Psychology Bulletin*, *12*(1), 51-62. doi: 10.1177/0146167286121006

Jones, S. M., Brown, J. L., & Lawrence Aber, J. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development*, *82*(2), 533-554. doi: 10.1111/j.1467-8624.2010.01560.x

Kahneman, D., & Tversky, A. (1982) The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201-208). New York: Cambridge University Press.

Kahneman, D., & Varey, C. A. (1990). Propensities and counterfactuals: The loser that almost won. *Journal of Personality and Social Psychology*, *59*(6), 1101-1110. doi: 10.1037/0022-3514.59.6.1101

Kestenbaum, R., & Gelman, S. A. (1995). Preschool children's identification and understanding of mixed emotions. *Cognitive Development*, *10*(3), 443-458. doi: 10.1016/0885-2014(95)90006-3

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42. doi: 10.1016/S0010-0277(02)00004-5

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*(8), 1134-1140. doi: 10.1177/0956797610376652

Lagattuta, K. H. (2005). When you shouldn't do what you want to do: Young children's understanding of desires, rules, and emotions. *Child Development*, *76*(3), 713-733. doi: 10.1111/j.1467-8624.2005.00873.x

Lagattuta, K. H. (2008). Young children's knowledge about the influence of thoughts on emotions in rule situations. *Developmental Science*, *11*(6), 809-818. doi: 10.1111/j.1467-7687.2008.00727.x

Lara, K. H., Lagattuta, K. H., & Kramer, H. J. (2019). Is there a downside to anticipating the upside? Children's and adults' reasoning about how prior expectations shape future emotions. *Child Development*, *90*(4), 1170-1184. doi: 10.1111/cdev.12994

Lecce, S., Caputi, M., & Hughes, C. (2011). Does sensitivity to criticism mediate the relationship between theory of mind and academic achievement?. *Journal of Experimental Child Psychology*, *110*(3), 313-331. doi: 10.1016/j.jecp.2011.04.011

LoBue, V., Nishida, T., Chiong, C., DeLoache, J. S., & Haidt, J. (2011). When getting something good is bad: Even three-year-olds react to inequality. *Social Development*, *20*(1), 154-170. doi: 10.1111/j.1467-9507.2009.00560.x

Loomes, G., & Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *The Review of Economic Studies*, 53(2), 271-282. doi: 10.2307/2297651

Ma, L., & Xu, F. (2011). Young children's use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition*, *120*(3), 403-411. doi: 10.1016/j.cognition.2011.02.003

MacLaren, R., & Olson, D. (1993). Trick or treat: Children's understanding of surprise. *Cognitive Development*, *8*(1), 27-46. doi: 10.1016/0885-2014(93)90003-N

Maguire, R., Maguire, P., & Keane, M. T. (2011). Making sense of surprise: An investigation of the factors influencing surprise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 176-186. doi: 10.1037/a0021609

Markman, K. D., & Tetlock, P. E. (2000). Accountability and close-call counterfactuals: The loser who nearly won and the winner who nearly lost. *Personality and Social Psychology Bulletin*, *26*(10), 1213-1224. doi: 10.1177/0146167200262004

McCloy, R., & Strange, P. (2009). Children’s understanding of counterfactual alternatives. In Proceedings of the 31st Annual Meeting of the Cognitive Science Society.

McMullen, M. N., & Markman, K. D. (2002). Affective impact of close counterfactuals: Implications of possible futures for possible pasts. *Journal of Experimental Social Psychology*, *38*(1), 64-70. doi: 10.1006/jesp.2001.1482

Medvec, V. H., Madey, S. F., & Gilovich, T. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology*, *69*(4), 603-610. doi: 10.1037/0022-3514.69.4.603

Medvec, V. H., & Savitsky, K. (1997). When doing better means feeling worse: The effects of categorical cutoff points on counterfactual thinking and satisfaction. *Journal of Personality and Social Psychology*, *72*(6), 1284-1296. doi: 10.1037/0022-3514.72.6.1284

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science*, 8(6), 423-429. doi: 10.1111/j.1467-9280.1997.tb00455.x

McDowell, D. J., O'Neil, R., & Parke, R. D. (2000). Display rule application in a disappointing situation and children's emotional reactivity: Relations with social competence. *Merrill-Palmer Quarterly*, *46*(2), 306-324.

Meyer, W. U., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion*, *21*(3), 251-274. doi: 10.1023/A:1024422330338

Nyhout, A., & Ganea, P. A. (2019). Mature counterfactual reasoning in 4-and 5-year-olds. *Cognition*, *183*, 57-66. doi: 10.1016/j.cognition.2018.10.027

O'Connor, E., McCormack, T., & Feeney, A. (2012). The development of regret. *Journal of Experimental Child Psychology*, *111*(1), 120-127. doi: 10.1016/j.jecp.2011.07.002

Ong, D. C., Goodman, N. D., & Zaki, J. (2015). Near-misses sting even when they are uncontrollable. In Proceedings of the 37th Annual Meeting of the Cognitive Science Society, 1775-1780.

Ong, D. C., Zaki, J., & Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition*, *143*, 141-162. doi: 10.1016/j.cognition.2015.06.010

Ong, D. C., Zaki, J., & Goodman, N. D. (2019). Computational models of emotion inference in theory of mind: A review and roadmap. *Topics in Cognitive Science*, *11*(2), 338-357. doi: 10.1111/tops.12371

Pesowski, M. L., & Friedman, O. (2015). Preschoolers and toddlers use ownership to predict basic emotions. *Emotion*, *15*(1), 104-108. doi: 10.1037/emo0000027

Peters, E., Slovic, P., Hibbard, J. H., & Tusler, M. (2006). Why worry? Worry, risk perceptions, and willingness to act to reduce medical errors. *Health Psychology*, *25*(2), 144-152. doi: 10.1037/0278-6133.25.2.144

Rieffe, C., & De Rooij, M. (2012). The longitudinal relationship between emotion awareness and internalising symptoms during late childhood. *European Child & Adolescent Psychiatry*, *21*(6), 349-356. doi: 10.1007/s00787-012-0267-8

Rieffe, C., Oosterveld, P., Miers, A. C., Terwogt, M. M., & Ly, V. (2008). Emotion awareness and internalising symptoms in children and adolescents: The Emotion Awareness

Questionnaire revised. *Personality and Individual Differences*, *45*(8), 756-761. doi: 10.1016/j.paid.2008.08.001

Rieffe, C., Terwogt, M. M., & Cowan, R. (2005). Children's understanding of mental states as causes of emotions. *Infant and Child Development*, *14*(3), 259-272. doi: 10.1002/icd.391

Rieffe, C., Terwogt, M. M., Petrides, K. V., Cowan, R., Miers, A. C., & Tolland, A. (2007). Psychometric properties of the Emotion Awareness Questionnaire for children. *Personality and Individual Differences*, *43*(1), 95-105. doi: 10.1016/j.paid.2006.11.015

Roese, N. J., & Olson, J. M. (Eds.). (2014). *What might have been: The social psychology of counterfactual thinking*. Psychology Press.

Ronfard, S., & Harris, P. L. (2014). When will Little Red Riding Hood become scared? Children's attribution of mental states to a story character. *Developmental Psychology*, *50*(1), 283-292. doi: 10.1037/a0032970

Ruffman, T., & Keenan, T. R. (1996). The belief-based emotion of surprise: The case for a lag in understanding relative to false belief. *Developmental Psychology, 32*(1), 40-49. doi: 10.1037/0012-1649.32.1.40

Russell, J. A. (1990). The preschooler's understanding of the causes and consequences of emotion. *Child Development*, *61*(6), 1872-1881. doi: 10.2307/1130843

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928. doi: 10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52. doi: 10.1016/S0010-0277(98)00075-4

Schank, R. C., & Abelson, R. P. (1975). Scripts, plans, and knowledge. In *Proceedings of the 4th International Joint Conference on Artificial Intelligence* (pp. 151-157). Morgan Kaufmann Publishers Inc.

Scott, R. M. (2017). Surprise! 20-month-old infants understand the emotional consequences of false beliefs. *Cognition*, *159*, 33-47. doi: 10.1016/j.cognition.2016.11.005

Shaw, A., & Olson, K. (2015). Whose idea is it anyway? The importance of reputation in acknowledgement. *Developmental Science*, *18*(3), 502-509. doi: 10.1111/desc.12234

Shepperd, J. A., & McNulty, J. K. (2002). The affective consequences of expected and unexpected outcomes. *Psychological Science*, 13(1), 85-88. doi: 10.1111/1467-9280.00416

Skerry, A. E., & Spelke, E. S. (2014). Preverbal infants identify emotional reactions that are incongruent with goal outcomes. *Cognition*, *130*(2), 204-216. doi: 10.1016/j.cognition.2013.11.002

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*(3), 303-333. doi: 10.1207/s15516709cog2803_1

Sunstein, C. R. (2002). Probability neglect: Emotions, worst cases, and law. *The Yale Law Journal*, *112*(1), 61-107. doi: 10.2307/1562234

Sunstein, C. R. (2003). Terrorism and probability neglect. *Journal of Risk and Uncertainty*, *26*(2-3), 121-136. doi: 10.1023/A:1024111006336

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, *104*(48), 19156-19159. doi: 10.1073/pnas.0700271104

Teigen, K. H. (1995). How good is good luck? The role of counterfactual thinking in the perception of lucky and unlucky events. *European Journal of Social Psychology*, *25*(3), 281-302. doi: 10.1002/ejsp.2420250304

Teigen, K. H. (1996). Luck: The art of a near miss. *Scandinavian Journal of Psychology*, *37*(2), 156-171. doi: 10.1111/j.1467-9450.1996.tb00648.x

Teigen, K. H. (1997). Luck, envy and gratitude: It could have been different. *Scandinavian Journal of Psychology*, *38*(4), 313-323. doi: 10.1111/1467-9450.00041

Teigen, K. H., & Keren, G. (2002). When are successes more surprising than failures?. *Cognition & Emotion*, *16*(2), 245-268. doi: 10.1080/02699930143000275

Teigen, K. H., & Keren, G. (2003). Surprises: low probabilities or high contrasts?. *Cognition*, *87*(2), 55-71. doi: 10.1016/s0010-0277(02)00201-9

Turnbull, W. (1981). Naive conceptions of free will and the deterministic paradox. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *13*(1), 1-13. doi: 10.1037/h0081115

van Dijk, W. W., & van der Pligt, J. (1997). The impact of probability and magnitude of outcome on disappointment and elation. *Organizational Behavior and Human Decision Processes*, 69(3), 277-284. doi: 10.1006/obhd.1997.2688

Weisberg, D. P., & Beck, S. R. (2010). Children’s thinking about their own and others’ regret and relief. *Journal of Experimental Child Psychology*, *106*(2-3), 184-191. doi: 10.1016/j.jecp.2010.02.005

Weisberg, D. P., & Beck, S. R. (2012). The development of children's regret and relief. *Cognition & Emotion*, *26*(5), 820-835. doi: 10.1080/02699931.2011.621933

Weisberg, D. S. & Gopnik, A. (2013). Pretense, counterfactuals, and Bayesian causal models: Why what is not real really matters. *Cognitive Science, 37*(7), 1368-1381. doi: 10.1111/cogs.12069

Wellman, H. M., & Banerjee, M. (1991). Mind and emotion: Children's understanding of the emotional consequences of beliefs and desires. *British Journal of Developmental Psychology*, *9*(2), 191-214. doi: 10.1111/j.2044-835X.1991.tb00871.x

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition, 30*(3), 239-277. doi: 10.1016/0010-0277(88)90021-2

Wellman, H. M., Kushnir, T., Xu, F., & Brink, K. A. (2016). Infants use statistical sampling to understand the psychological world. *Infancy, 21*(5), 668-676. doi:10.1111/infa.12131

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*(3), 245-275. doi: 10.1016/0010-0277(90)90024-E

Widen, S. C., & Russell, J. A. (2010). Children's scripts for social emotions: Causes and consequences are more central than are facial expressions. *British Journal of Developmental Psychology*, *28*(3), 565-581. doi: 10.1348/026151009X457550d

Widen, S. C., & Russell, J. A. (2010). Young children’s understanding of others’ emotions. In Lewis, M., Haviland-Jones, J. M., & Barrett, L. F. (Eds.), *Handbook of Emotions* (348-363). Guilford Press.

Widen, S. C., & Russell, J. A. (2011). In building a script for an emotion, do preschoolers add its cause before its behavior consequence?. *Social Development*, *20*(3), 471-485. doi: 10.1111/j.1467-9507.2010.00594.x

Wu, Y., Baker, C. L., Tenenbaum, J. B., & Schulz, L. E. (2018). Rational inference of beliefs and desires from emotional expressions. *Cognitive Science*, *42*(3), 850-884. doi: 10.1111/cogs.12548

Wu, Y., & Schulz, L. E. (2018). Inferring beliefs and desires from emotional reactions to anticipated and observed events. *Child Development*, *89*(2), 649-662. doi: 10.1111/cdev.12759

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112*(1), 97-104. doi: 10.1016/j.cognition.2009.04.006

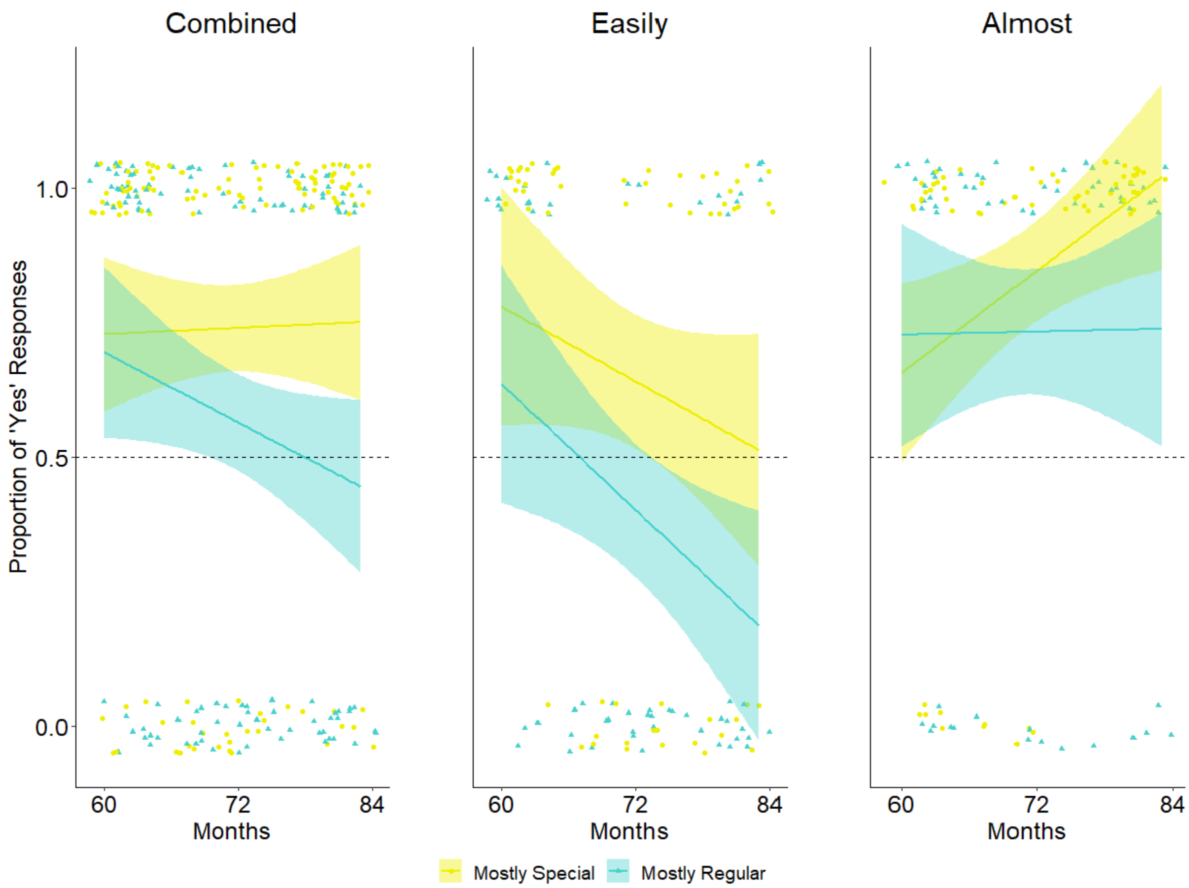
Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012-5015. doi: 10.1073/pnas.0704450105

Xu, F., & Tenenbaum, J. B. (2007a). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, *10*(3), 288-297. doi: 10.1111/j.1467-7687.2007.00590.x

Xu, F., & Tenenbaum, J. B. (2007b). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245-272. doi: 10.1037/0033-295X.114.2.245

Appendix A

Graph of close counterfactual responses split by question-type for Experiment 3 (Chapter Four)



Appendix B

Analyses according to pre-registration for Experiment 4 (Chapter Four)

In our pre-registration, we reported that we would run the analysis for the happiness ratings using a paired-samples t-test (we report GEEs in the main text because they are more appropriate for this ordinal data). Consistent with the GEE, adults rated the character as happier about receiving a regular balloon in the Mostly Regular condition than in the Mostly Special condition, $t(226) = 9.38$, $p < .001$. We also reported that in order to look at how adults' responses to the counterfactual questions corresponded with their happiness ratings, we would conduct a 2X2 ANOVA with distribution as the within-subjects factor and pattern-type as the between-subjects factor. As with the GEE, we found a main effect of distribution, $F(1,225) = 38.38$, $p < .001$, a main effect of pattern-type, $F(1,225) = 13.84$, $p < .001$, and a distribution-by-pattern-type interaction, $F(1,225) = 13.47$, $p < .001$. We followed-up with t-tests. Participants who responded both "yes" in the Mostly Special condition and "no" in the Mostly Regular condition rated the character as being happier about receiving a regular balloon in the Mostly Regular condition than in the Mostly Special condition, $t(173) = 10.68$, $p < .001$. Participants who responded to the counterfactual questions in any other way showed no differences in happiness ratings between the two conditions, $t(52) = 1.27$, $p = .209$.

Appendix C

Chapter Two New Analyses

Experiment 1

A Generalized Estimating Equations (GEE) binary logistic regression with age in months mean-centred and entered as a covariate revealed a significant main effect of age, *Wald* $X^2(1) = 10.68$, $p = .001$. Pairwise comparisons revealed that 4-year-olds gave significantly fewer correct responses compared to children at all other ages, *p*s ≤ .019. However, responses did not significantly differ between 5-, 6-, and 7-year-olds, *p*s ≥ .214. Tests against chance revealed that 7-year-olds predominantly chose the correct character (*M* = 1.43, *SD* = .817), *p* =.010, 6-year-olds showed a trend in this direction (*M* = 1.27, *SD* = .785), *p* =.072, 5-year-olds responded at chance (*M* = 1.17, *SD* = .874), *p* =.297, and 4-year-olds predominantly chose the wrong character (*M* = 0.67, *SD* = .802), *p* =.032. These analyses are consistent with the original analyses in which age was binned into years.

Experiment 2

A GEE binary logistic regression with age in months mean-centred and entered as a covariate and condition (impossible, improbable) as a between-subject factor revealed a marginal effect of age, *Wald* $X^2(1) = 3.63$, $p = .057$. There was no effect of condition, *Wald* $X^2(1) = 2.23$, $p = .135$, and no age by condition interaction, *Wald* $X^2(1) = 0.21$, $p = .647$. Pairwise comparisons revealed that 6-year-olds performed significantly better than 4-year-olds, *p* =.006, and 5-year-olds, *p* =.038, but 4-year-olds and 5-year-olds did not differ in their performance, *p* = .635. Tests against chance revealed that 6-year-olds (*M* = 1.33, *SD* = .764) predominantly chose the correct character, *p* =.011, whereas, 4-year-olds (*M* = 0.90, *SD* = .632), *p* = .314, and 5-year-olds (*M* = 0.98, *SD* = .768), *p* = .835, performed at chance. In the original analysis, when age was binned into years, there was a significant main effect of age, however, it is now marginal. Follow-up tests revealed that the findings are consistent with the original analyses.

Experiment 3

A GEE binary logistic regression with age in months mean-centred and entered as a covariate and condition (belief, probability, control) as a between-subject factor revealed a significant effect of age, *Wald* $X^2(1) = 4.63$, $p = .031$ (6-year-olds chose the correct character more often than 5-year-olds), a marginally significant effect of condition, *Wald* $X^2(2) = 5.44$, $p = .066$, and no age by condition interaction, *Wald* $X^2(2) = 4.21$, $p = .122$. Pairwise comparisons revealed that children in the probability prompt condition performed significantly better than children in the control prompt condition, $p = .012$, but no differences were found between children's performance in the belief and control prompt conditions, $p = .796$. Children's performance in the probability prompt condition was also significantly better than their performance in the belief prompt condition, $p = .040$. Tests against chance revealed that only 6-year-olds performed above chance levels, $p = .003$. In the original analyses when age was binned into years, all main effects and interaction were marginal. The lack of interaction in these new analyses suggest that 5- and 6-year-olds might not differ much in their ratings between conditions.

Experiment 4

The analyses for this experiment would be the same as reported in Chapter Two as I only tested one age range (6-year-olds) so age would not be entered into the model. However, I also analyzed this data with age in months in case there were developmental differences within the span of one year. A GEE binary logistic regression with age in months mean-centred and entered as a covariate and condition (belief, probability, control) as a between-subject factor revealed a significant effect of condition, *Wald* $X^2(2) = 9.46$, $p = .009$, no effect of age, *Wald* $X^2(1) = 1.43$, $p = .232$, and no age by condition interaction, *Wald* $X^2(2) = 0.21$, $p = .901$. Pairwise comparisons revealed that children in the probability prompt condition (*M* = 1.20, *SD* = .951), attributed surprise to the character significantly more than children in the control prompt condition (*M* = 0.35, *SD* = .671), $p = .001$, but no differences were found between children's performance in the belief (*M* = 0.65, *SD* = .875) and control prompt conditions, $p = .212$. Children in the probability prompt condition also attributed surprise to the character marginally more than children in the belief prompt condition, $p = .051$.Tests against chance revealed that only children in the probability prompt condition chose surprise at above chance levels, with chance being 0.50 out of 2, as there were four possible emotions to choose from and two trials, $p = .007$. These new findings are consistent with the original findings.

Chapter Three New Analyses

Experiment 1

A GEE ordinal logistic regression with age in months mean-centred and entered as a covariate and condition (Mostly Yummy, Mostly Yucky) as a between-subject factor revealed a significant age by condition interaction, *Wald* $X^2(1) = 4.86$, $p = .027$. There was a marginal effect of condition, *Wald* $X^2(1) = 3.66$, $p = .056$ (children gave higher ratings in the Mostly Yucky condition than in the Mostly Yummy condition), and no main effect of age, *Wald* $X^2(1) = 2.31$, $p = .129$. I explored each age group separately and found that both 5- and 6-year-olds rated the girl as significantly happier in the Mostly Yucky condition than in the Mostly Yummy condition, *Wald* $X^2(1) = 4.01$, $p = .045$, and, *Wald* $X^2(1) = 8.97$, $p = .003$, respectively. However, 4-year-olds did not show differences between the two conditions, *Wald* $X^2(1) = 1.27$, $p = .260$. These findings are consistent with the original ANOVA.

Experiment 2

A GEE ordinal logistic regression with age in months mean-centred and entered as a covariate and condition (Mostly Yummy, Mostly Yucky, 50/50) as a between-subject factor revealed a main effect of condition, *Wald* $X^2(2) = 10.49$, $p = .005$, and a main effect of age, *Wald* $X^2(1) = 4.07$, $p = .044$. However, looking at the effect of age, 5- and 6-year-olds did not differ in their ratings, $p = .098$. There was no age by condition interaction, *Wald* $X^2(2) = 0.11$, $p = .948$. Follow-up tests revealed that children rated the girl as significantly happier in the Mostly Yucky condition than in the Mostly Yummy condition, $p = .003$, and happier in the 50/50 condition than in the Mostly Yummy condition, $p = .035$. Happiness ratings did not significantly differ between the Mostly Yucky and 50/50 conditions, $p = .620$. The original ANOVA did not reveal a main effect of age, however following up on the new analysis suggest that there is no difference between ages. Thus, the findings from the new analyses are consistent with the original analyses.

Experiment 3

A GEE ordinal logistic regression with age in months mean-centred and entered as a covariate and condition (Mostly Yummy, Mostly Yucky) as a between-subject factor revealed a main effect of condition, *Wald* $X^2(1) = 15.60$, $p < .001$, where children rated the outcome as better in the Mostly Yucky condition than in the Mostly Yummy condition. There was no effect of age, *Wald* $X^2(1) = 2.75$, $p = .097$, and no age by condition interaction, *Wald* $X^2(1) = 0.001$, $p = .972$. These findings are consistent with the original findings.

Experiment 4

A GEE ordinal logistic regression with condition (Mostly Yummy, Mostly Yucky) as a within-subject factor and judgment-type (Happiness, Quality) as a between-subject factor revealed a condition by judgment-type interaction, *Wald* $X^2(1) = 6.99$, $p = .008$. There were no main effects of condition, *Wald* $X^2(1) = 2.60$, $p = .107$, or judgment-type, *Wald* $X^2(1) = 0.57$, $p = .452$. When rating quality, children gave higher ratings in the Mostly Yucky condition than in the Mostly Yummy condition, $p = .002$. However, when judging happiness, there was no significant difference between the Mostly Yummy and Mostly Yucky conditions, $p = .551$. These findings are consistent with the original findings.