The development of integrating number and proportion in probabilistic decision making

by

Reem Tawfik

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, 2021

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	Kristy vanMarle Associate Professor Psychological Sciences University of Missouri
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	John Turri Professor Department of Philosophy University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

The ability to integrate multiple pieces of information and use it to guide decision-making is an essential part of everyday reasoning. While it may not seem like it, the information in our environments is often numerical in nature. From simple decisions like which cashier line to stand in at the grocery store, to more consequential judgments like evaluating the chances of getting accepted into a competitive graduate program, numbers and proportions are everywhere. And while combining numerical information to make judgments and decisions sounds challenging, even young children have some of the requisite abilities to do so. In this dissertation, I describe and discuss a series of experiments that examine the developmental trajectories of integrating numerical and proportional information to make probability judgements and arrive at favourable outcomes in game-like scenarios. Chapter 2 examines whether 5- and 6-year-old children (N=160) and adults (N=68) can integrate two types of numerical information to make decisions in a probability game involving single- and multi-draw samples from different distributions. I presented children with a computer game in which they must maximize the number of green objects obtained. In order to do so successfully they were required to integrate the absolute number of draws with the proportion of targets to non-targets from which those draws are made. Across five studies, I established that 5- and 6-year-olds and adults can – under certain conditions –integrate two sources of numerical information to make decisions that maximize the odds of a favourable outcome. Chapter 3 examines the developmental origins of these integrative abilities by adapting the paradigm used in Chapter 2 to test infants (*n*=46). I presented them with two trial types: one where the correct response was to choose the lower draw number from the distribution with a higher proportion of target objects, and another where the correct response was to choose the higher draw number from the

distribution with a lower proportion of target objects. Results from 10-12-month-olds revealed that infants were not able to combine the numerical and proportional information to make probability judgments and performed at chance levels with no effects of age or trial type. These results suggested that the ability to integrate both numerical and proportional information to make probability judgments is not yet a consistent hallmark of probabilistic reasoning within the age range we tested. Chapter 4 examines whether or not toddlers (n=40) would be successful on the same trial types used with infants. Results from 18-30-month-olds show that toddlers were able to correctly choose the larger draw number from the distribution with the lower proportion of target objects, but responded at rates no different than chance on the trial where the correct response was to choose the smaller draw number from the distribution with the larger proportion of target objects. Taken together, the results from these three sets of experiments suggest that adults and school-aged children are capable of integrating probability and number in a game-like probability task, and that these abilities may begin in toddlerhood. It appears that beginning in the second year of life, but likely not before then, human learners can consider both the distributions and the number of draws when reasoning about sampling and probability.

Acknowledgements

First and foremost, I would like to thank my supervisor Stephanie Denison. Without you, this work would not be possible; thank you for believing in me, supporting me, and providing the right balance between motivation and room to explore and grow. I would also like to thank the other faculty members that have supported and inspired me on this journey.

My fellow graduate students (past and present), lab-mates, and friends: I could not have made it through these past few years without your moral support. You are all incredible and being a small part of all of your amazing journeys has been a true pleasure. Words cannot convey the impact your kindness has had on me; thank you.

I want to acknowledge and thank Tiffany Doan and Madeline Crichton for the role they played in data collection, and Julianna Lu for her invaluable help with data visualization.

To my incredible sisters and best friends, Rawan and Hana: I love you both so much. Knowing that I had your unconditional support meant the world to me. Thank you for the babysitting, the pep talks, and most importantly, the fun & hysterical laughter throughout all of this. It meant more than you will ever know. Mom and dad, I honestly would not even be here if it weren't for you. This accomplishment is the result of many things, most importantly of which is your hard work and sacrifice these past 3 decades. I can write a whole dissertation on how I owe this all to you. I love you both so incredibly much.

Hisham, you have been the most supportive and incredible partner ever. When we got married a year into this program, I had no idea just how important of a role you would play in me actually making it to the finish line. You were there for me on all those long nights and meltdowns, and you truly helped me not just get through this academically, but also to keep up with the million other things in life. You are the most amazing husband, father, and friend.

Dedication

This dissertation is dedicated to my children. You are the true meaning of happiness and fulfilment. I hope that one day you can actually read this and be proud of your mom.

Table of Contents

List of Figures

List of Tables

Chapter One: General Introduction

After a long day at work, you get into your car and assess your options. It is just before rush hour, so you consider taking the short, but typically more-congested, route home as opposed to the longer, but usually much less-busy, option. Time is ticking and traffic is (likely) building, and you have to make this decision under uncertainty, knowing that it could cost you precious minutes (if not hours) that can be better spent relaxing at home. Individuals frequently engage in this kind of decision making in everyday life and in contexts where they might hardly notice that the available information requires a complex decision based on their intuitions about quantities. In this dissertation, I discuss and investigate a particular numerical challenge that has not been explored before, at least in early childhood: the ability to combine two sources of numerical information in probabilistic decision making, and how this ability develops and changes across the lifespan. In particular, I am interested in how information about small numbers can be combined with estimates of ratios of large numbers to compute and compare probabilities of different outcomes.

While this appears challenging, we know that humans have expectations about numbers from birth, and that they have some of the pre-requisite abilities that may provide the foundation to solve these kinds of problems by approximately 6 months. In this chapter, I will review some of the literature on infant, child, and adult numerical cognition, paying special attention to small number and large number representations, as well as review the relevant work on proportional and probabilistic reasoning. I will then outline the (relatively sparse) literature on children's abilities to combine multiple sources of information and discuss the objectives and methods of the research I conducted for this dissertation.

The Development of Numerical Reasoning Abilities

Reasoning about small numbers

How do infants, children, and adults keep track of small numbers of items or objects? Many agree that tracking small numbers of individual items (typically four or fewer) and holding representations of their features in memory requires the use of a system often referred to as parallel individuation (Choo & Franconeri, 2014; Feigenson & Carey, 2003, 2005; Feigenson et al., 2002; Feigenson et al., 2004; Lipton & Spelke, 2004; Revkin et al., 2008; Simon, 1997; Uller et al.,1999; Xu, 2003). The consensus is that this system represents individual objects, capturing numerical information indirectly via those representations, and that the limit on its capacity comes from the number of objects it can represent in parallel (Kahneman et al., 1992; Pylyshyn, 1989, 1994). There is a large body of work showing that parallel individuation is available quite early in infancy (e.g., Feigenson et al., 2002; Hyde & Spelke, 2011). In a now classic task examining this ability, infants were presented with choices between the number of crackers placed in opaque buckets. When presented with small collections of crackers, 10-12-month-olds chose the bucket with the larger number of crackers when there were no more than three in each bucket (Feigenson et al., 2002; vanMarle, 2013). They also succeeded at choosing the larger of two quantities when comparing only numbers that are larger (e.g. 4 vs. 8); but interestingly, performed at chance when one of the buckets in the comparison contained more than three and the other contained three or fewer (e.g., 1 vs. 4, 2 vs. 4, 3 vs. 6, 2 vs. 8), despite including controls to rule out alternative explanations such as presentation length and attentiveness (Feigenson et al., 2002; vanMarle, 2013). This failure to make these latter comparisons, coupled with success at making comparisons with three or fewer crackers is taken as evidence that numerosities of 3 and under are reasoned about using a different mechanism than that which is used for numerosities of 4 or larger in infancy.

In addition to representing small numbers of objects and allowing a comparison of quantities held in memory, parallel individuation appears to support other numerical reasoning in infancy (Carey, 2009; Feigenson & Carey, 2003, 2005; Le Corre & Carey, 2007, 2008). For example, Feigenson and Carey (2003) showed that infants likely represent hidden objects with a unique symbol that is held in working memory, subsequently allowing them to search for and retrieve hidden objects. When they retrieve the number of objects for which they have a running tally, infants stop searching. It also appears that infants can use parallel individuation to represent more than one collection at a time, employing chunking to group together up to two or three chunks of up to three objects, which would allow for a reorganized representation of these individual items into a larger set (Moher et al., 2012; Zosh et al., 2011). This allows them to keep track of, and subsequently be able to compare, two separate collections with a maximum capacity of two objects in one collection and three in the other (Feigenson, et al., 2002; Feigenson & Halberda, 2008; Rosenberg & Feigenson, 2013). They can then proceed to compare these sets to determine equivalence, or lack thereof. By the preschool years, children are adept at comparing sets of small numbers, with very good accuracy (Cheung & LeCorre, 2018; Margolis, 2020)

Further, infants can discriminate between numbers in the small number range even when continuous variables such as surface area are controlled, which suggests that detecting change in set numbers relies on *number* as opposed to other low-level perceptual features that may be correlated with number (Feigenson, 2005; Feigenson & Carey, 2003). To test whether infants actually track numerosity or whether they make judgements based on continuous characteristics such as area, Cordes and Brannon (2009a) habituated 7-month-olds to two or three same-size squares and then showed half of the infants the old number with a new contour and the other half the old contour with the new number at test. Results showed that infants responded to both numerical changes as well as contour changes. How infants compared the mental models for the test trial stimuli with those of the habituation stimuli can explain these results. In the new contour/old number condition, they compared the sum of the continuous properties that were bound to the visual stimuli, causing them to notice the change in contour length. In the old contour/new number condition, they compared the models for one-to-one correspondence, allowing them to notice that they no longer match in numerosity (Cordes & Brannon, 2009a; Margolis, 2020).

The parallel individuation system allows for numerical representations and comparisons, but it also allows for somewhat complex computations and operations. One would never expect that infants under one year of age can engage in arithmetic, but astonishingly there is evidence that very young infants have the ability to track objects and form accurate expectations about the outcomes of making changes to small numbers of objects. When 5-month-olds are tested in a violation of expectation looking-time paradigm by being shown a character that joins another character behind a screen on stage, infants look significantly longer when the screen is removed to reveal only one character as opposed to two (Wynn, 1992). The same pattern of results is found with subtraction; when infants are shown two characters on stage prior to a screen occluding their view, and observe one of those characters walking off-screen, they appear to be surprised when the screen is removed to reveal two characters as opposed to one. This updating of objects and set numerosity was interpreted as indicating that infants can form expectations about transformations to small number sets. To rule out other low-level expectations such as physical location expectations (i.e. the infants could have simply memorized where on stage various characters should be, and when this is violated it could have led to the same pattern of results) this result was replicated in a rotating location condition (Koechlin, et al., 1997). Even when controlling for potentially confounding location-based expectations by making the positions of the objects unpredictable, 5-month-olds still looked longer at arithmetically impossible events than possible ones. One additional piece of evidence for infants having a rudimentary ability to compute small number arithmetic come from audio-visual studies that show that when 5-month-olds hear a particular number of sounds associated with falling objects, and a screen is removed to reveal fewer visual objects than the sounds indicated, they looked significantly longer (Kobayshi et al., 2004; 2005). Taken together, these findings indicate that even young infants can track objects and have some expectations about the outcomes of these simple changes involving small numbers (e.g. 1+1; 1+1+1; 2-1).

However, there appears to be a discontinuity between small numbers (up to a limit of four) and large numbers in the spontaneous processing of numerosities, in particular in the infant stage. Some experiments show that infants can discriminate all numerosities in the range of one to four, but fail to discriminate pairs that include a numerosity outside that range (e.g., 3 & 6), even when, as in the example, their ratio is greater than the ratio between discriminable pairs of four or less (Feigenson, et al., 2002; vanMarle, 2013). How can we reconcile these latter findings with the finding that infants do discriminate 1:2 ratios with pairs such as 8 & 16 (Lipton & Spelke, 2003; Xu & Spelke, 2000)? One account that explains the underlying mechanisms that would lead to these seemingly discrepant findings is that representations involving large numbers utilize a different system altogether: the approximate number system (ANS). Evidence for the existence of a separate system comes not only from the different discrimination abilities mentioned previously, but also from physiological markers. For example, event-related potentials (ERPs) display neural signatures that are unique to either the small or the large number systems (Hyde & Spelke, 2011). I now turn to the evidence for the system that allows for representing and reasoning about numbers that lie outside of the small number range: the approximate number system.

Reasoning about large numbers

A large body of literature makes a compelling case for a system of numerical representation that nonhuman animals, human infants, children and adults all share for representing approximate numerosities (Dehaene, 1997/2011, Feigenson et al., 2004; Gallistel & Gelman, 2000; 2005; Mix et al., 2002). This approximate number system (ANS) allows for a quick estimate of a large number of objects, regardless of modality, when they are presented too briefly for counting and when other variables such as area, perimeter, density, and so forth are controlled for (Barth et al., 2003; Lipton & Spelke, 2003, 2004; Xu & Spelke, 2000; Dehaene, 2009). These numerical representations are approximate, and they obey Weber's Law such that the ability to discriminate two numerosities is dictated by the ratio of the two quantities, not the absolute numerical difference, and this becomes more precise with age (Halberda & Feigenson, 2008; Lipton & Spelke, 2003; 2004; Odic et al., 2013; Whalen et al., 1999; Xu & Spelke, 2000). There are also physiological markers of this ratio-dependent number discrimination in infancy, as findings show that EEG-measured activity differs as a function of the ratio between a habituated and novel numerosity in 7-month-old infants (Libertus & Brannon, 2009).

This ability to discriminate large numbers extends to comparing two sets or arrays of objects, and the acuity with which these judgements are carried out does not depend on the total number of objects in the array, but rather on the ratio between the objects presented (Lipton &

Spelke, 2004; Xu, 2003). Whereas adults discriminate numerosities at about a 1.15 ratio (Van Oeffelen & Vos, 1982), 6-month-old infants discriminate numerosities that differ by a factor of 2.0 (4 vs. 8 and 8. vs. 16 elements, but not 4 vs. 6 or 8 vs. 12), and 9-month-old infants discriminate numerosities that differ by a factor of 1.5 (4 vs. 6 and 8. vs. 12, but not 4 vs. 5 or 8 vs. 10) (Cordes & Brannon, 2008; Lipton & Spelke, 2003, 2004; Wood & Spelke, 2005; Xu, 2003; Xu & Spelke, 2000). Even newborn infants are able to discriminate between dot arrays that differ by a ratio of 1:3 (e.g. 6 vs. 18), and cross-modal studies have shown that they are capable of matching numerical quantities across sets even when they were presented with a number of items auditorily and tested with visual numerosity (Izard et al., 2009). In this study, newborns looked longer at visual displays that contained the same number of elements as the number of tones they were familiarized with, suggesting that infants possess a capacity to abstract numerosity from birth.

Another important finding that lends support to the notion that infants do indeed track approximate numerosities is that infants as young as 6 months of age can discriminate numerosities of that differ by a 2:1 ratio in visual displays, even when carefully controlling for potential confounds such as area covered by the stimuli, stimuli size, and density of stimuli (Xu & Spelke, 2000). After habituating to a visual array of dots in a particular numerosity in the large number range (for example 16 dots), infants looked significantly longer at subsequent test trials where they were presented with a visual array of a numerosity that differed by a 2:1 ratio (8 dots or 32 dots). Strikingly, this effect was found in spite of controlling for potentially confounding variables such as dot brightness, density, and size, indicating that the discrimination is indeed occurring on the basis of number. This ratio-dependence also extends beyond discrimination to comparisons of “more” and “less”. When 10- to 12-month-olds are prompted to choose between two sets of hidden food items, they succeeded at choosing the larger number only when the sets differed by a ratio of 2:3 (e.g. 6v9) but not when the sets differed by a ratio of 3:4 (6v8) or 7:8 (7v8) (vanMarle et al., 2016). This converging evidence suggests that the ANS not only supports large number representations and discrimination, but also ordinal relations in infancy.

Taking this a step further, there is also some evidence that infants as young as 11 months understand the ordinal relations between numbers, and are not simply sensitive to differences between numerosities in the large number range. When infants were habituated to a series of ascending or descending sequences of three numerical displays (e.g. 4-8-16 or 16-8-4), 11-month-olds (but not 9-month-olds) dishabituated when they were shown numerical displays in the opposite direction, even when different sets of numbers where displayed in the two directions (Brannon, 2002). Understanding ordinal relations has also been widely studied beyond infancy. Preschoolers are able to tell which of two visual displays has “more” objects, and are just as accurate at these judgements cross-modally (i.e. auditory to visual stimuli; Barth et al., 2005). This suggests that these ordinal judgements are not based on lower-level modality-specific factors. Accuracy on these tasks also adhered to a ratio-dependent sensitivity threshold that implicates the ANS. Finally, the ANS operates throughout the lifespan and is still present in adulthood. Although the precision of the system changes over time, the signature properties of the system remain intact throughout development: The system continues to operate abstractly, transcending modalities, with numbers represented approximately and imprecisely (Barth et al., 2003).

In addition to supporting ordinal relations, the ANS also supports more complex computations, such as addition, subtraction, multiplication, and division, starting in the infant months. Infants can perform approximate, non-symbolic arithmetic and have correct expectations about the outcomes of these computations. When 9-month-olds are shown the addition of an array of 5 objects to another array of 5 objects, they look longer at the unexpected outcome of 5 than the expected outcome of 10, indicating that they already have expectations about the approximate results of these calculations (McCrink & Wynn, 2004). The same pattern of results was found in conditions that involved subtraction. Notably, these effects are found despite researchers controlling for other variables such as total area covered by the objects and contour length, suggesting that the infants' judgements and expectations are rooted in discrete numerosities. Similar success was found in children as young as 5 years of age engaging in non-symbolic multiplication and division computations, and having correct (though of course, approximate) expectations about the products and quotients of these computations despite having no formal education in these areas (McCrink & Spelke, 2010; 2016). Once again, accuracy on these problems increased as a function of the ratio between numerosities, a hallmark of the ANS.

Adults also show similar patterns of performance on tasks that required the addition, subtraction, multiplication, and division of nonverbally estimated numerosities, where formal arithmetic operations are not used. Subjects were shown two numerosities such that they were displayed too briefly for them to be counted. Then, they were shown a third numerosity, and they indicated by pressing one of two buttons whether the sum, or difference, or product, or quotient of the first two numerosities was greater or less than the third. The numerosities were presented either as dot arrays (controlling for dot density and overall area covered) or as tone sequences. In some conditions, presentation modalities were mixed, so for example, subjects compared the sum of a tone sequence and a dot array to either another tone sequence or another dot array. Results showed that there was no effect of the magnitude of the numerosities on accuracy or reaction time, but there was, an effect found for the ratios between numerosities (Barth et al., 2006). Adults' performance on these tasks involving non-symbolic approximate computations also display the signature ratio-sensitivity associated with the ANS.

Reasoning about proportions

In addition to being able to represent sets of items (both small and large) infants, children, and adults also have a basic ability to reason about ratios and proportions. From a relatively young age, children show an ability to discriminate and match based on proportionality, as opposed to absolute quantity or size. For example, understanding that a container with 30 green lollipops and 30 black lollipops is proportionally-equivalent to a container with 10 green lollipops and 10 black ones, despite the overall number of green lollipops being different, involves an understanding of the relationship of 30 to 30, and 10 to 10, and subsequently being able to compare these two relations to determine equality (or lack thereof). Children starting at the age of 3 can indicate with of two shapes has a proportionally larger target area and match figures that have the same ratio of discrete coloured areas (Boyer & Levine, 2015; Boyer et al., 2008; Hurst & Cordes, 2018). Additionally, toddlers as young as 2 years of age can match the size of a target wooden peg when it is presented inside of a container by using the proportional information of peg length to container height, whereas they had greater difficulty in matching the size of the pegs alone in the absence of the container as a referent (Hutternlocher et al., 2002).

Some studies that investigate children's proportional reasoning use a map-reading task. For example, when 3- and 4-year-olds are given a map with a rectangle on it that represents a sandbox and a dot marking where in the sandbox a toy was buried, most children were able to carry out the proportional translation of map to actual sandbox and find the missing toy (Huttenlocher et al., 1999). This paradigm is perhaps less relevant in the context of this dissertation, but it demonstrates that children as young as 3 can engage in proportional reasoning along at least one dimension (in the case of the sandbox map, length). Another paradigm that is frequently used in studies that investigate proportional reasoning is a proportional matching task. In these tasks, children are given a picture of a shape that is composed of different-coloured segments and asked to match it to another shape that is scaled up or down, or even displayed in a different orientation. Preschoolers can match shapes that share proportional information on the basis of height-width ratios (of rectangles, in this case) as well as different coloured segments despite transformations of size and orientation (Sophian, 2000). Similarly, when shown a pizza-shaped figure that is split up into eighths and removing a particular number of “slices,” children are able to remove the same proportion of a different pizza that is segmented into quarters (Singer-Freeman & Goswami, 2001).

Research on proportional reasoning in children has highlighted some of its most important elements. First, many of the failures reported in studies of proportional reasoning in slightly older children (usually children 6 years and older) can be attributed to tracking exact quantities of target objects as opposed to representing the relation of target objects to non-target objects, or target objects to overall number of objects. There is a thorough discussion of how this impacts performance on proportional reasoning tasks that present continuous variables as opposed to discrete ones, but this lies outside of the scope of this dissertation (Boyer et al., 2008; Boyer & Levine, 2015; Mix et al., 2002). Another interesting feature of proportional reasoning using large quantities of objects and/or uncountable features that would likely require the use of the ANS is that the inherent imprecision involved in this kind of reasoning leads to a reliance on intuitive, as opposed to quantitative, judgements. These intuitions are the result of an informal reasoning system that does not require or invoke effortful calculation (Kahneman & Tversky, 1982; Tversky & Kahneman, 1974). In the absence of exact numerical information, individuals have no other option but to rely on less precise intuitions to approximate proportions. In a temperature mixing task, 10-12-year-olds and 14-15-year-olds were given information about two liquids of different temperatures and asked what the resulting temperature would be of a new mixture combining them. The information that was provided was either intuitive or quantitative in nature. The intuitive condition involved showing participants pictures and descriptions of the quantities and the temperatures of the two original liquids. In the quantitative condition the exact numerical values of the amounts and temperatures of the original liquids was provided, and participants were instructed to "use math." Results showed that the order of the tasks had a significant influence on accuracy in the numerical tasks, such that getting the intuitive information first improved performance on the numerical tasks. The same effect was not found for performance on the intuitive tasks (Ahl et al., 1992).

Work with adults has also suggested that the ANS is involved in representing proportions among whole numbers. When college students were asked to choose the larger fraction when symbolic arrays were pitted against discrete non-symbolic arrays (visually displayed as dot arrays for the numerator and the denominator), and again when these discrete non-symbolic arrays were pitted against continuous non-symbolic arrays in which the numerator and denominator were each replaced by a circle of variable area, they succeeded in these comparisons even though they answered too quickly to have explicitly counted the dots in the discrete non-symbolic arrays (Matthews and Chesney, 2015). Additionally, their reaction times and errors displayed the ratio-dependent signature of the ANS.

Work directly investigating proportional reasoning or ratio-based reasoning in infancy is very sparse. The only relevant work on representing ratios and proportions in infancy investigates young infants' ability to discriminate between ratios using the ANS. This work suggests that infants as young as 6 months of age are capable of comparing different proportions, with a sensitivity threshold similar to that of discrimination of large numerosities. In the only published study to my knowledge investigating this ability, McCrink and Wynn (2007) habituated 6-month-olds to a particular ratio (for example, 1:4) of blue circles to yellow pacman shapes by showing them multiple examples of that ratio (independent of the individual number of elements on each habituation trial). At test, they were presented with alternating trials, both differing in the absolute numbers of items seen on the habituation trials but one displaying another instance of the original 1:4 ratio and the other displaying a different, 1:2 ratio. Six-month-olds dishabituated to the different ratio, indicating that they were able to discriminate between ratios that differed by a factor of 2 (though they did not dishabituate in another condition when the ratios differed by a factor of 1.5). Interestingly, this dishabituation occurred despite careful control of continuous variables of the stimuli, such as area, density, and total overall number. This mirrors the discrimination threshold that Xu and Spelke (2000) found where 6-month-olds could reliably discriminate absolute numerical values that differed by a 2:1 ratio but not a 3:2 ratio, which points to the involvement of the ANS in computing the proportions between large numerosities as well as the proportions between proportions.

The Development of Probabilistic Decision Making

Just as numerosity is inherent in many human experiences from birth, statistics and probability are also present. Infants appear to track transitional probabilities among syllables as a potential mechanism for early language learning (Aslin et al., 1998; Saffran et al., 1996) and they also have basic intuitions about probability that allow them to make large generalizations from small amounts of data (see Bryant & Nunes, 2012; Denison & Xu, 2019 for a review). For the purpose of this dissertation, I will only review the sampling literature. 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). Infants as young as 6 months of age (though not 4 months) reliably look longer at outcomes that violate their expectation (as shown by studies using the violation-of-expectation looking time paradigm) when viewing sampling events. If a container has mostly pink balls and a few yellow balls, and an experimenter randomly samples 4 yellow balls and 1 pink ball, infants look longer at that outcome, indicating that it is unexpected or surprising, as compared to a random sample of 4 pink balls and 1 yellow ball, indicating lack of surprise at that outcome (Denison et al., 2013). This suggests that infants have a basic understanding that the sampled objects should match the proportional characteristics of the population from which the objects were sampled. The opposite is also true; they expect a population to be proportionally-congruent to sampled objects that come from that population (Xu & Garcia, 2008).

Ten- to 12-month-old infants also demonstrate that probabilistic reasoning abilities can guide their choices and actions. In a two-alternative forced choice paradigm, infants were shown two transparent containers that have target and non-target lollipop-type items (coloured objects attached to sticks) in different ratios, and the experimenter sampled from these containers by placing a lollipop from each jar in an opaque cup without allowing the infant to see its identity. The infants reliably walked or crawled towards the sample that had come from the jar with the higher ratio of their preferred lollipop. Crucially, this persisted even when the jar with the higher ratio of target-to-non-target lollipops contained fewer preferred lollipops in absolute number (Denison & Xu, 2014). In fact, even non-human animals like great apes and capuchins display a rudimentary understanding of statistics and probabilities based on proportional information.

They are able to choose a sample taken from a population with a higher probability of yielding a preferred item (Drucker et al., 2016; Eckert et al., 2017; 2018; Rakoczy et al., 2014; Tecwyn et al., 2017).

The work on probabilistic competence between infancy and the age of five suggests that 3- and 4-year-olds sometimes struggle with problems that require the use of posterior evidence, proportions or combinatorial procedures (e.g., Girotto & Gonzalez, 2008; Gonzalez & Girotto, 2011; Piaget & Inhelder, 1975; Siegler, 1981), as well as seemingly simple tasks in which they only have to consider prior information in order to predict an uncertain event. For example, given a bag containing 3 yellow chips and 1 blue chip, 4-year-olds answer at chance level, if they have to predict whether a randomly drawn chip will be yellow or blue (Girotto & Gonzalez, 2008). Likewise, 3-year-olds perform at chance level if they have to predict whether a ball, bouncing inside a rectangular box with one hole on one side and three holes on the opposite side, will exit from the one-hole side or from the three-hole one (Teglas et al., 2007 – Experiment 3). Children in this age range succeed under deterministic conditions, but fail when having to deal with proportional information that requires evaluating competing probabilities (Girotto et al., 2016). As an example of this, researchers presented children (aged 3-5 years) with a set of target tokens and a set of non-target ones, and randomly sampled one object from each set. They then asked children to choose which sample they want to get the target token. Children understood the task, as evidenced by their success on the deterministic trials, where one set contained only attractive tokens and the other one contained only unattractive tokens. However, on the trials that required forming probabilistic expectations, 4- and 5-year-olds made optimal choices. Only the youngest children performed randomly and/or were guided by superficial heuristics, like preferring the set containing a larger number of attractive tokens rather than the set containing a greater proportion of attractive tokens (Girotto et al., 2016). Examining the gap wherein 3-year-olds appear to be unable to engage in some (even very simple) explicit probability-based tasks while infants and non-human animals have succeeded is outside the scope of the present dissertation. In the experiments presented here, I will test 5- to 6-year-old children in an explicit verbal task, and children under 3 years in a non-verbal action task.

Probabilistic reasoning in older children using games of chance sheds light on their understanding of proportions and sampling, and how these in turn influence the odds of particular outcomes. For example, when 6- to 10-year-olds were presented with pairs of spinners that were divided into different coloured sections and told that if a spinner landed on one colours they would win stickers, children were able to correctly choose which spinner had a better chance of winning based on the proportions of coloured segments (Jeong et al., 2007). These probabilistic judgements that rely on proportional information also appear to be affected by the ratio of proportions, such that a judgement involving two competing distributions that are further apart is easier than one where the competing distributions are closer together. In other words, accuracy correlates with the ratio of proportions. As an example, 6- and 7-year-old children were presented with a game where they had to choose which of two of marble collections to sample in order to get a marble of a target colour. Each marble collection contained two different colours of marbles, varying both in overall number as well as ratio of target to non-target colours. Of particular interest, the pairs of competing marble collections were carefully designed such that children saw a wide range of ratios between each of the proportions, starting with the easiest comparison of 70% vs. 5% all the way to the most challenging comparison of 55% vs. 50%. As the ratio of proportions between the competing marble collections became larger, accuracy increased (O'Grady & Xu, 2020). The same kind of ratio-dependent performance on probabilistic judgments involving proportional information is found in studies with adults (O'Grady et al., 2016).

Combining Multiple Sources of Information in Decision Making

Going back to the example of choosing which route to take on your commute home, it is perhaps now clear to see how integrating multiple pieces of numerical information involved in this scenario is an integral part of making optimal decisions. In the comparative literature, there is extensive work establishing nonhuman animals' ability to combine various sources of quantitative information, such as number of rewards and time, to make decisions about different rates of return (Balci et al., 2009; Gallistel & Gibbon, 2000). I will review the literature on children's abilities to combine multiple sources of information in more detail in Chapter 2, as most of this work examines these abilities in school-aged children, which is the age range of interest in that chapter. This literature sheds some light on children's numerical integration in probabilistic decision making. One particularly relevant line of work focuses on investigating or children's ability to make expected value judgments based on integrating the probability and value of a winning outcome, as well as making estimates of task difficulty in game-like settings (Bayless & Schlottmann, 2010; Schlottmann, 2001; Schlottmann & Anderson, 1994; Schlottmann & Tring, 2005). This work reveals that even children as young as 4 years of age are able to integrate multiple factors when making these sophisticated judgements. In a similar vein, 5-year-old children can use numerical information to update an already formed probability judgment when given an additional piece of information that alters the sampling space (Girotto & Gonzalez, 2008). And finally, children as young as 4 years can evaluate and choose between two sources of numerical information in a judgment task when the two pieces of information cue opposite inferences (Gualtieri et al., 2020).

Overview of Dissertation

This dissertation sheds light on whether or not we can combine reasoning about small and large numbers in a way that allows us to effectively judge probabilities of different outcomes. In Chapters 2-4 of this dissertation, I describe and discuss a series of experiments that aim to shed light on the ability of infants, toddlers, school-aged children, and adults to integrate number and proportion in probabilistic decision making. To do so, I adapted the lollipop task (Denison & Xu, 2014) to introduce both numerical and proportional information in a game-like manner that then allows participants to make judgements based on odds. The task itself consists of a game during which the goal is to maximize chances of getting target objects. The set-up always consists of two large containers of target and non-target lollipops. These containers are displayed to participants such that they can obtain an estimate of the ratio of items in the containers, and so that they cannot count up the exact number of objects. The experimenter then draws a number of objects from the container, seemingly randomly, and places them in separate cups adjacent to the large containers. The exact number of lollipops taken from each container is displayed to the participants for the duration of the task, represented by the lollipop sticks emerging from the cups (with the identity – target or non-target – not visible to participants). The goal of this task is to use the information gathered from the proportions in the containers, as well as the number of objects in the sample (each signifying a potential opportunity to get a target-coloured object), to choose the sample cup that would maximize their chances of getting target objects.

These probabilities can be construed of (and thus calculated) in a number of ways. The question of *exactly how* participants in chapters 2 – 4 estimate and compare these probabilities will be left open. Below I will outline two ways that these odds could be calculated.

Figure 1 provides a visual depiction of all types of trials included in the dissertation. Throughout the dissertation, the trial type will be written with the shorthand seen here, where the proportion of target items in each container is written first and the respective number of draws second. As an example, in the 100:0 1v3 trial, this corresponds to the percent of green (targets) in container 1 (100%) and the percent of green in container 2 (0%) with a single item drawn from the 100% container and three items drawn from the 0% container. The choice that would result in better odds of getting targets is always coloured in red font in the trial type shorthand. Note that in Figure 1, targets are depicted as green lollipops and non-targets as black, as was the case in the Chapter 2 stimuli. In Chapters 3 and 4, the targets were pink (non-targets were still black).

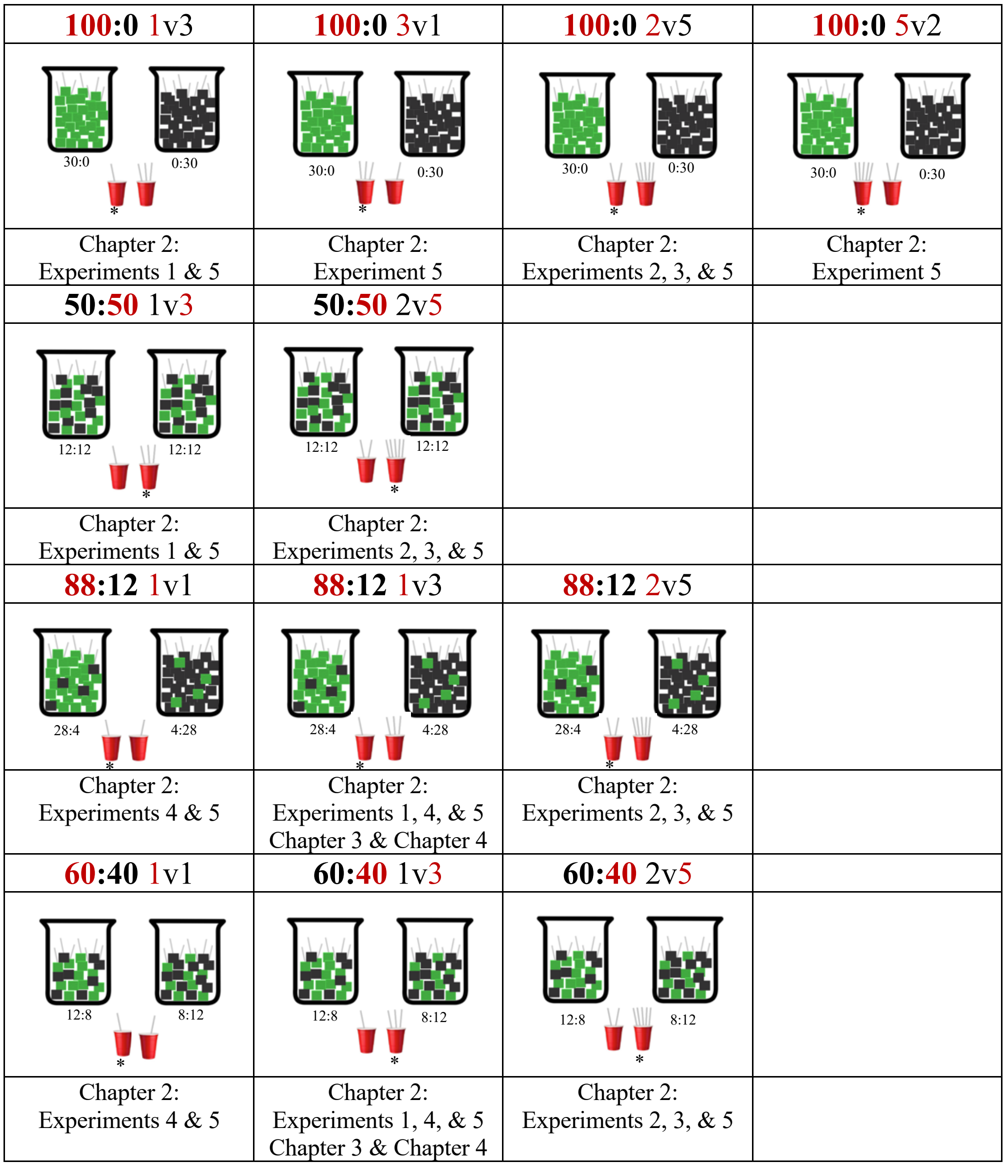

Figure 1. Summary of all trials included in this dissertation and corresponding experiments in which each one appears.

The top row contains entirely straightforward cases, where the best choice is absolutely clear: always choose the sample that was drawn from the all-green container, no matter how many items are drawn from either container. These types of trials can establish whether a participant of a particular age can make the correct choice based entirely on the contents of the containers, without being incorrectly swayed by the number of draws (three is more than one, but this does not matter when the one will certainly be green and the three will certainly all be black).

Row 2 depicts a second set of straightforward cases where the proportions are equal (50% green) in each container, but the number of draws differ. In these cases, participants should base their choices on the number of draws, always choosing the higher draw to increase the number of green lollipops they might get.

Row 3 depicts a set of more challenging problems, some of which require integration of the proportions with the number of draws to arrive at the correct choice, or the choice that maximizes the chances of getting targets. In the first two trials in this row, the container with 88% green is so advantageous on the ratio dimension, that a larger number of draws does not tip the scale far enough to make the larger number of draws from the 12% green container the better choice. That is, the participant has better odds of ending up with some amount of green going with the smaller number of draws from the 88% container. The first trial in this row compares single draws from each container, which is a more straightforward trial that examines whether participants choose the correct container based only on the proportions.

Row 4 also depicts some of the more challenging (mathematically, the most challenging) trials. In the second two trials in this row, while the 60% green container is more advantageous than the 40% green container, this is a small difference. Thus, the larger number of draws from the less advantageous containers does tip the scale just far enough to make those the better choice. These trials were the most difficult to design. I was aiming to create a complementary set of trials to those in row 3. In those trials, the participants must choose the smaller number of draws and the better ratio to arrive at the best choice. In the trials in row 4, I was aiming to create cases where the participants must choose the worse ratio with the better number of draws. These are difficult to create though, because the ratios must be distinct enough such that the difference is detectable (if they are any closer to 50:50, it would be questionable as to whether the participants detect the difference between the ratios). Thus, it is simply difficult to pull these trials very far apart mathematically without risking making the ratios indistinguishable or having to make a problematic number of draws from the 40% container (i.e., increasing the number of draws too much makes this work impossible to conduct with infants due to the small number comparison limits). The first trial in this row compares single draws from each container. This type of trial is informative because it examines whether participants detect the difference in ratios between the two containers (participants should choose the 60% container on a 1 versus 1 draw).

Computing objectively correct choices for the trials

One possible computation that a participant could engage in when doing these tasks is to determine the odds of getting "at least one" target item in each of the sample containers and comparing those odds. I consider this a potential strategy for children, because it is hard to know exactly how the older children (Chapter 2) interpret the instructions and how the infants and toddlers (Chapters 3 and 4) might conceive of the task, given that they receive no explicit instructions. A second possibility, which I see as a very likely potential strategy, due to the instruction to "get as many green ones [targets] as you can" to older children and adults, is to choose the sample cup that is likely to have the highest number of greens. While both computations favour the same choice for all trials, they are mathematically different. I present these probabilities in the Appendix.

Row 1 is straightforward (the 100% container always gives a 100% chance at green and the largest number of green no matter how the participant conceives of the goal).

Row 2: For the 50:50 1v3 trial, the odds of obtaining at least 1 target is 50% versus 89%, respectively. In the 50:50 2v5 trial, the odds of obtaining at least 1 target is 76% versus 98%, respectively. For computations that maximize the number of greens obtained, participants could consider how likely different draws are from each of the distributions and the likelihood of them occurring together. Then one can consider the odds of "winning" (getting more green ones) by choosing one cup over the other. For example, in 50:50 1v3 there are two possible outcomes for the single draw and four possible outcomes for the three-item draw (see Table 1 for the probabilities). Based on this, a person is 11 times more likely to win by choosing the three-item sample: there is one scenario in which a person wins with the single draw, which has a 6.25% chance of occurring and five scenarios in which they win by choosing the three-item draw, which have a combined 69% chance of occurring (there are two ties, which we ignore). In 50:50 2v5, one is 12.3 times more likely to win by choosing the 3-item draw (Table 2).

Row 3: For 88:12 1v3, the odds of obtaining at least 1 target is 88% versus 34% and for 88:12 2v5, the odds of obtaining at least 1 target is 99% versus 51%. For 88:12 1v3, one is 8.2 times more likely to obtain more greens when they choose the single-item draw (Table 3) and for 88:12 2v5, one is 19 times more likely to obtain more greens when choosing the two-item draw (Table 4). For 88:12 1v1, it is simply 88% versus 12%.

Row 4: For 60:40 1v3, the odds of obtaining at least 1 target is 60% versus 81% and for 60:40 2v5, the odds of obtaining at least 1 target is 85% versus 95%. For 60:40 1v3, one is 4 times more likely to obtain more greens when choosing the three-item sample (Table 5) and for 60:40 2v5, one is 3 times more likely to obtain more greens when choosing the five-item sample (Table 6). As can be seen from these calculations, the 60:40 trials (with either the 1v3 or the 2v5 draws) are quite close together mathematically, which does make them particularly challenging.

For 60:40 1v1, they simply compare 60% and 40%.

Chapter Two: Integration of Number and Proportion in Children and Adults

The ability to integrate multiple pieces of information and use it to guide decision making is an essential part of everyday reasoning. While it may not seem like it, the information in our environments is often numerical in nature. From simple decisions like which cashier line to stand in at the grocery store, to more consequential judgments like evaluating the chances of getting accepted into a competitive graduate program, numbers and proportions are everywhere. And while using numerical information, particularly proportional information, to make judgments and decisions sounds challenging, even young children have some of the requisite abilities to do so. In the current experiments I investigate the use of numerical and proportional information in a single probabilistic reasoning task in school-aged children and adults.

Though the literature on children's numerical integration in probabilistic decision making is sparse, there are at least two relevant lines of research that include children in the age range of interest (5-6 years). The first of these examines whether or not children, like adults, are able to make expected value judgments based on integrating both the probability and value of a winning outcome in a game (e.g., Schlottmann, 2001; Schlottmann & Anderson, 1994; Schlottmann & Tring, 2005). Expected value is defined as the product of the probability of a particular outcome and the value of that outcome, or EV = P x V. For example, in a game of chance where you had a 50% probability of winning, and the value of the prize was quantified as 10 tokens, the expected value of playing this game would be 5. Alternatively, if a similar game also presented players with a 50% chance of winning but the prize was only worth 5 tokens, the expected value of this game would be 2.5. When presented with a roulette game, 5- to 10–year-olds were asked to judge how happy a puppet would be to play a game in which the probability of winning the game and the value of the prize varied on each trial. They found that even the youngest children were able to integrate both factors when making these third-person judgements, but the oldest children used more mature integrative strategies than the youngest children. That is, only the older children relied on a multiplicative strategy that is more robust and provides the highest expected value in challenging cases (Schlottmann & Anderson, 1994).

Further, children as young as 5 years of age are able to use normative multiplicative reasoning when evaluating probability in a more intuitive context (Bayless & Schlottmann, 2010). In a marble rolling game, in which the objective was to roll a marble into a target goal, two difficulty dimensions were manipulated: size of the target and distance between the target and the rolling position. This created a probability judgement that relies on estimates of skill and objective task difficulty – represented as continuous dimensions, which, under some circumstances, are easier for children to reason about (e.g., Boyer et al., 2008). Five- and seven-year-olds were asked to make difficulty judgments for a variety of trials that differed in these dimensions prior to actually playing the game themselves. They were also asked to rate how happy they would be to play the game when the difficulty and the value of a prize were varied across games. Both age groups responded normatively in a multiplicative fashion (Bayless & Schlottmann, 2010). They accurately integrated information on both the size of the target and its distance when making difficulty judgements. They also integrated information on the difficulty of the game and the value of the prize when determining how happy they would be to play the game, realizing that a game where the distance of the marble to the target was shortest, the size of the target was largest, and the value of the prize was highest, was most desirable.

In addition to being able to effectively combine multiple pieces of quantitative information to judge task difficulty and desirability of a particular outcome, children as young as 5 years of age are also able to update estimates of posterior probabilities based on new information that changes the sampling space (Girotto & Gonzalez, 2008). When presented with 4 circle chips (all black) and 4 square chips (1 black and 3 white), and asked to make a prediction about the colour of a chip drawn at random, children predominantly responded correctly by saying ‘black.’ However, when the researcher sampled a chip, kept it hidden in their hand and said that it felt like a square, most children then correctly responded by saying they thought it would be white. This implies that not only are children capable of using information from different sources effectively to make decisions about events, they can also use new information as it arises in order update posterior probabilities. Taken together, these findings lend support to the idea that young children are able to integrate multiple pieces of information to make judgments about probabilistic tasks.

This past research has looked at young children’s abilities to integrate probability and other numerical information, or two sources of information where only one is numerical in nature. In the present experiments, I examine whether 5- to 6-year-old children can integrate two sources of numerical information to make decisions about the probability of an unknown outcome. Specifically we present children with a game in which they must maximize the number of green objects obtained. In order to do so successfully they will be required to integrate the absolute number of blind draws with the proportion of targets to non-targets, from which those draws are made. Five- and 6-year-old children (N=160) were presented with two containers of green and black lollipops, with the goal of getting as many green lollipops as possible, in a series of game-like trials. We manipulated the proportions of coloured lollipops in the containers, as well as the number of items drawn from each container across various trials. For example, in one type of trial, a single (unknown) lollipop was drawn from a very high proportion of green:black lollipops and three (unknown) lollipops were drawn from a very low proportion of green:black lollipops.

When children are asked to choose which of two samples would yield a more favourable outcome (i.e. give them a better chance of getting green lollipops) we expected that they might employ one of three strategies: First, they might focus exclusively on the dimension of absolute number, using only the information from the number of sampled items. In this case, they would always opt for the sample with the higher number of lollipops, disregarding the proportion of objects from which those samples were drawn. Children this age are capable of discriminating between the numbers of draws in our experiments (i.e., 1 vs 3 in Experiments 1 and 4; and 2 versus 5 in Experiments 2 and 3; Cheung & LeCorre, 2018), thus children should be capable of evaluating choices on the basis of this dimension.

Second, they might favour the proportional information and always choose the draw from the container with the largest proportion of desired objects, regardless of the absolute number of draws sampled from it. We know that children this age should be capable of this because 4- and 5-year-old children do this readily in two-alternative forced-choice tasks (Denison et al., 2006; Girotto et al., 2016; Yost et al., 1962), and even young infants (Denison & Xu, 2010, 2014) as well as non-human animals (Rakoczy et al., 2014; Tecwyn et al., 2017) are capable of this.

The third potential strategy would be an integrative one where children should adapt their strategy based on both numerical information (absolute number of draws) as well as proportional information (ratio of desirable to undesirable objects). It is currently unknown as to whether children are able to combine information in this way to maximize outcomes. In the following experiments, I explore these open questions in an attempt to better understand these potentially competing strategies and their behavioural manifestations in school-aged children as well as adults.

Experiment 1

Methods

Participants

Data from 40 5- and 6-year-old children were included in the final analyses (*Mean age*: 6;1, *range*: 5;0 – 6;11, *females:* 19). Sample sizes were decided in advance, based on a lab stopping rule, but no formal power analyses were conducted. The choice of 40 participants was based on previous work (e.g., Gualtieri et al., 2020). An additional 3 children were tested and their data were excluded for failure to complete the experiment (*n*=1); and an inability to indicate a clear choice for just one cup (*n*=2). Children were tested individually in schools and museums in the Waterloo Region and a museum in the Greater Toronto Area.

Procedure

All stimuli were displayed on a laptop using PowerPoint slides. Children were shown a green lollipop and a black lollipop on one slide. They were told they would play a game in which the goal was to get as many green lollipops as possible. They were then asked a comprehension question to make sure that they understood the goal of the game (e.g. "how do you win the game?"). We had a planned inclusion criterion that children were required to answer this question correctly, but no children answered this incorrectly.

Following this, there were four trials. The general procedure of each trial was as follows (see Figure 2.1 for a schematic): A slide presented two large transparent containers holding different distributions of green (target) and black (non-target) lollipops. Then two opaque but empty cups appeared in front of the containers. The containers were covered by opaque grey boxes such that the distributions of lollipops were no longer visible. The experimenter told the participant that someone will pick some lollipops without looking at the colours inside the containers. An animated hand then reached into each of the containers (one at a time) and drew 1 lollipop from one container and 3 lollipops from the other (drawing the sample of three, one at a time). The hand picked up each lollipop in a way such that only the stick was visible, not the colour of the lollipop; this was also true of the way the lollipops were placed in the cups, only the sticks were visible. When both containers had been sampled, the grey covers disappeared to remind participants of the containers' contents. The experimenter then asked the participant "which cup would you pick to get green?" and asked them to point towards the cup of their choice. This question was intended to be ambiguous grammatically, using "green" as a mass noun, to avoid signalling to the child whether they should choose the cup with multiple lollipops or a single lollipop. However, this may have been interpreted to mean "green" in the singular sense, but I return to this in the discussion section.

Design

Each child saw four trial types, in counterbalanced order, that varied the distribution of lollipops in the containers and the number of draws taken from each container (see Figure 2.2).

In the 100:0 1v3 trial, there was one container with 30 green lollipops and no black lollipops, and another with 30 blacks and no greens. The hand drew 3 lollipops from the all-black container and 1 from the all-green container. In this case, children should choose the 1-item sample.

In the 50:50 1v3 trial, there were identical distributions in the two containers (12 green and 12 black) and the hand drew 3 lollipops from one container and 1 from the other. In this case, children should choose the 3-item sample.

These first two trial types are included to ensure that children will make correct choices on the basis of the individual dimensions and to ensure that they follow the game instructions. Thus, in the 100:0 1v3 trial, the number of draws is varied and the simplest possible proportions (0 vs. 100%) are used to ensure children can use the proportion dimension, without errantly focusing on the number of draws. In the 50:50 1v3 trial, the proportions are held constant and the number of draws is manipulated, using a comparison that we know children this age should be sensitive to (Cheung & LeCorre, 2018) to ensure children can use number of items in the sample when the proportions are equated.

The 88:12 1v3 and the 60:40 1v3 trials test children's integration of these two dimensions. In the 88:12 1v3 trial, there was one container with 28 green and 4 black lollipops and another with the opposite distribution (4 green and 28 black lollipops). The hand sampled 1 lollipop from the green-majority container (high proportion of target objects, low draw number) and 3 lollipops from the black-majority container (low proportion of target objects, high draw number). In this case, children should choose the 1-item sample taken from the high proportion container.

In the 60:40 1v3 trial there was one container with 8 green and 12 black lollipops and another with 12 green and 8 black lollipops. The hand sampled 1 lollipop from the green-majority container (high proportion of target objects, low draw number) and 3 lollipops from the black-majority container (low proportion of target objects, high draw number). In this case, children should choose the 3-item sample picked from the lower proportion container.

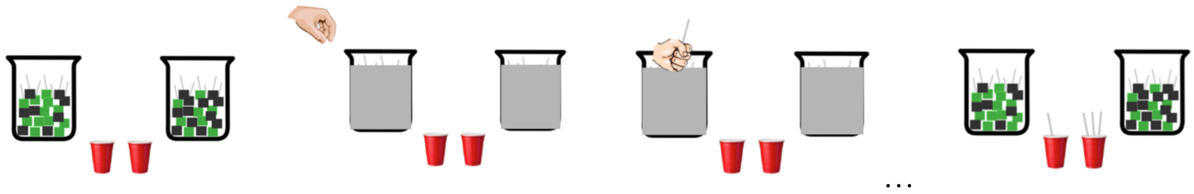

Figure 2. Schematic of the general procedure of each test trial

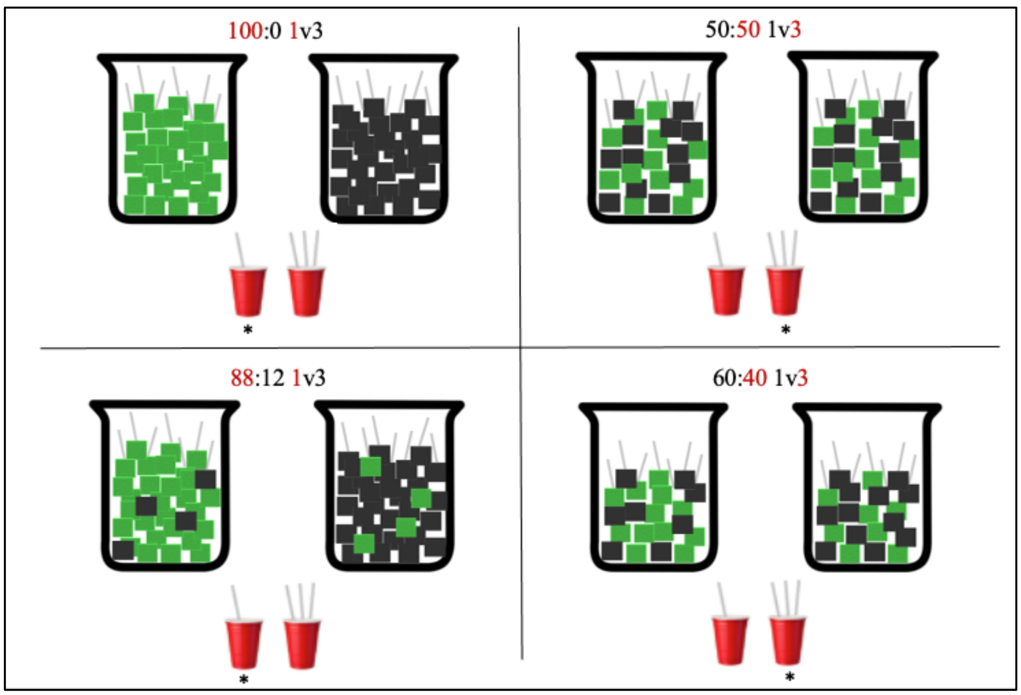

Figure 3 Depiction of the four trials in Experiment 1. Correct choices based on the computations in Chapter 1 are indicated by the asterisks.

<u>Results</u>

In all four experiments, all children chose the green lollipop following the comprehension check at the beginning of the experiment. We coded the data such that the draw that had a greater probability of yielding green was coded as "correct." This meant that a correct response on any given trial was coded as a 1 and an incorrect response was coded as a 0.

A Generalized Estimating Equations (GEE) binary logistic regression with trial as a between-subject factor and age (in months) centred around the mean as a covariate revealed a significant main effect of trial on accuracy, *Wald* $X^2(df = 3, N = 40) = 119.77$, $p < .0001$. Figure 4 shows the proportion of children producing correct responses on each trial. No effect of age, nor interaction of trial by age, was found. Figure 2.3 shows children's average correct responses on each trial. Bonferroni-corrected pairwise comparisons revealed that performance on the 60:40 1v3 trial was significantly poorer than the other three trials ($p < .0001$). Average accuracy on the 50:50 1v3 trial was also lower than both the 100:0 1v3 trial ($p < .0001$) as well as the 88:12 1v3 trial ($p = .002$).

To take a closer look at performance on each trial, I conducted binomial tests against chance for each trial after coding for accuracy. On the 100:0 1v3 trial, where the outcome was guaranteed, all children selected the 1-lollipop draw. On the 50:50 1v3 trial, 62.5% of children answered correctly, choosing the 3-lollipop draw. A binomial test indicated that this was not significantly different from chance (p=.154). On the 88:12 1v3 trial, children succeeded at choosing the draw that has a higher probability of yielding a green lollipop (M=92.5%, SD=.267, p<.0001) where choosing the 1-lollipop draw from the container with a higher proportion of green lollipops is correct. However, on the 60:40 1v3 trial, they showed a preference for the 1-lollipop draw from the container with a larger proportion of green lollipops, even though it yields a lower probability of getting a green lollipop than the 3-lollipop draw from the container with a smaller proportion of green lollipops. They performed significantly lower than chance on this last trial type (M=30%, SD=.464, p=.016).

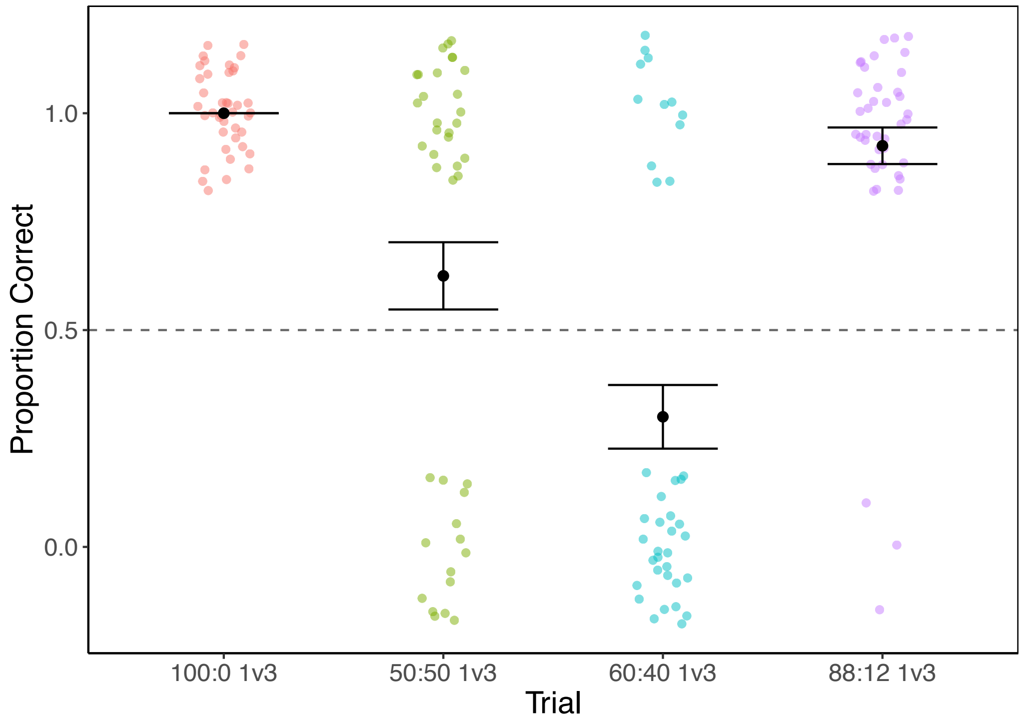

Figure 4. Mean correct responses on each trial indicated by dot (Experiment 1). Error bars indicate ± 1 standard error of the mean and individual data points are jittered for visibility. The dashed line represents chance (50%) performance.

<u>Discussion</u>

These results suggest that children are able to correctly choose which cup gives them a greater chance of obtaining green lollipops when the proportion of green lollipops in the target container was greater, regardless of draw number. On the 100:0 1v3 and the 88:12 1v3 trials, children correctly chose the 1-lolipop draw from the container with all or mostly green lollipops. However, on the 50:50 1v3 trial, children chose each sample at rates no different from chance, suggesting an inability to use the absolute number comparison even in this relatively easy type of trial. Finally, on the 60:40 1v3 trial where the 3-lollipop draw from the container with mostly *black* lollipops gives children a greater chance of getting green than the 1-lollipop draw from the container with mostly green lollipops, children chose the latter at greater than chance levels.

Overall, children's performance is mostly consistent with two potential interpretations: one is that they quite closely tracked the proportion of items in the containers, showing little sensitivity to number of draws. On each trial, children made selections as a group that favoured the proportions (even in the case of the 50:50 proportion, responses were statistically not different from chance, which would be predicted by a proportion-based strategy). An alternative interpretation though, is that children favoured the single lollipop draw. This is mostly consistent with their behaviour, with the exception of the 50:50 proportion, in which children chose the 3-item draw 63% of the time, not different from chance but leaning toward the multi-item draw, rather than the 1-item draw. Despite being intentionally vague in the wording of the task instructions to avoid implying that children should choose the sample container with multiple objects or one object (I instructed children to "get green"), it is still possible that my instructions inadvertently biased participants to choose the 1-lollipop draw on all trials. Specifically, it is possible that instructing them to pick the cup that would optimize their chances of "getting green" may have led children to interpret that as being "green" in the singular sense rather than our intended ambiguous meaning. In order to tease these two possibilities apart, in the next experiment I changed the stimuli to only include multiple draw numbers, removing the ambiguity surrounding the wording.

Experiment 2

<u>Methods</u>

Participants

Data from 40 5- and 6-year-olds were included in the final analyses for this experiment (*Mean age*: 6;1, *range*: 5;0 – 6;9, *females:* 14). Children were recruited from and individually tested in local schools and museums in the Waterloo Region and the Greater Toronto Area.

Procedure

The procedure was identical to Experiment 1 except for one change in the design. Instead of a 1-lollipop versus a 3-lollipop draw during each trial, this experiment presented a 2-lollipop versus a 5-lollipop draw. Proportions of target to non-target objects in the containers were identical to Experiment 1.

Design

The design was identical to Experiment 1. For example, in the trial type with uniform distributions of 30 green or 30 black in each container, the hand drew 5 lollipops from the black container and 2 from the green container. See Figure 5 for a schematic and the correct choices.

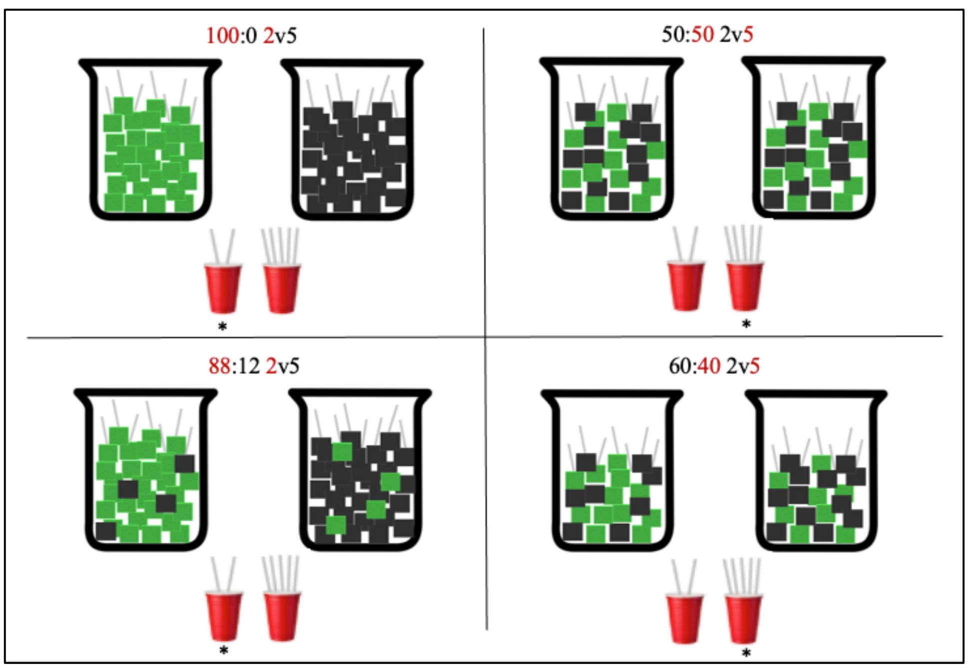

Figure 5. Depiction of the four trials in Experiments 2 & 3. Correct choices based on the computations in Chapter 1 are indicated by the asterisks.

Results

A GEE binary logistic regression with trial as a between-subject factor and age centred around the mean as a covariate revealed a significant main effect of trial on accuracy, *Wald* X^2(*df* = 3, $N = 40$) = 50.716, $p < .0001$. No effect of age, nor interaction of trial by age, was found.

Figure 2.5 shows the proportion of children producing correct responses on each trial. Bonferroni-corrected pairwise comparisons revealed that performance on the 60:40 2v5 trial was significantly poorer than the other three trials ($p < .0001$), and performance on the 100:0 2v5 trial was significantly better than performance on both the 50:50 2v5 and the 88:12 2v5 trials ($p = .038$).

I then conducted binomial tests against chance for each trial to follow up on these results. On the 100:0 2v5, where the outcome was guaranteed, all children selected the 2-lollipop draw. On the 50:50 2v5 trial, children responded correctly by choosing the 5-lollipop draw at a rate significantly greater than chance levels (M=90%, SD=.304, p<.0001). On the 88:12 2v5, children succeeded at choosing the 2-lollipop draw that has a higher probability of yielding a green lollipop (M=90%, p<.0001). Finally, on the 60:40 2v5, children's responses were not different from chance, as 48% of participants correctly chose the 5-lollipop draw (SD= .504, p=.636).

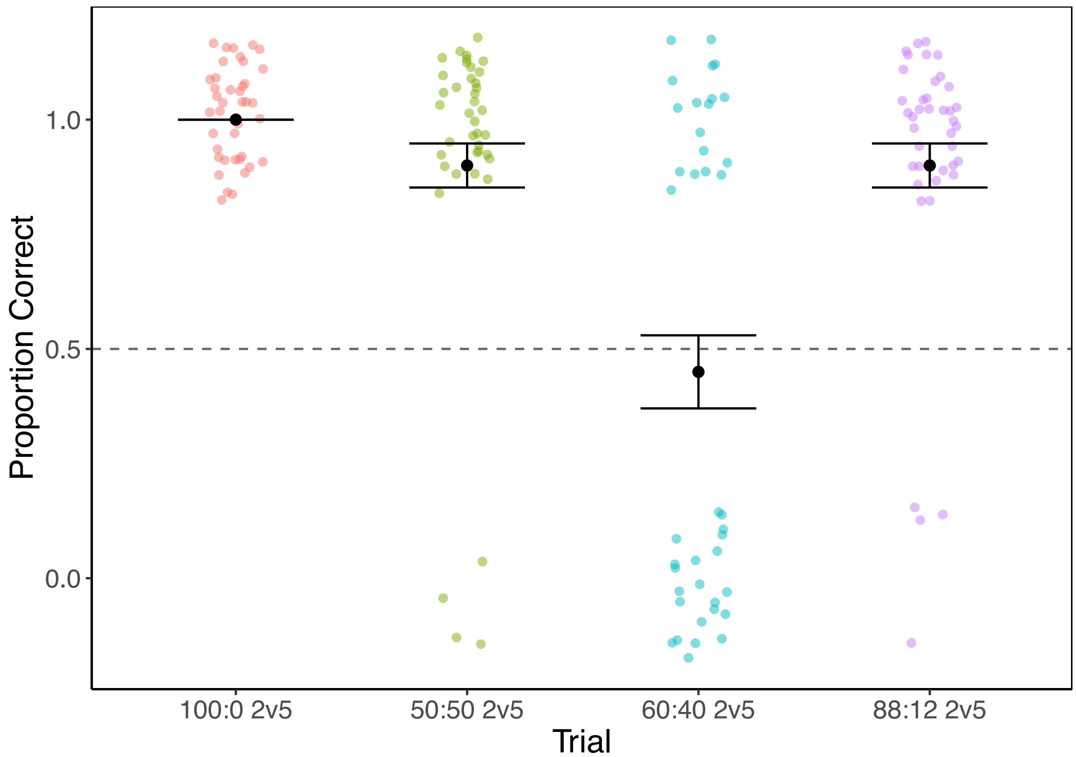

Figure 6. Mean correct responses on each trial indicated by dot (Experiment2). Error bars indicate ± 1 standard error of the mean and individual data points are jittered for visibility. The dashed line represents chance (50%) performance.

Discussion

Children's performance on the 100:0 2v5 and the 88:12 2v5 trials, where choosing the small draw number from the container with a higher proportion of green lollipops was more likely to yield green, is almost identical to Experiment 1. Performance on the 50:50 2v5 trial drastically improved in this experiment. Children correctly chose the 5-draw lollipop at greater than chance levels. This suggests that children struggled with the 1 versus 3 comparison, either because they were unable to make that judgment correctly based on comparing those numbers or because the experimental set-up somehow biased them in favour of choosing single-number draws. I lean toward the latter interpretation, given that previous literature suggests that children this age are aware that 3 is greater than 1 (Cheung & LeCorre, 2018). Although, notably, it cannot be assumed that just because children recognize that one set contains more items than another set, that they also recognize that having more draws in a lottery results in higher odds of winning. Therefore, this experiment is the first to my knowledge showing that children this age appreciate that increasing the number of draws increases the odds of winning in a random game.

Finally, while performance on the 60:40 2v5 trial was not different than chance, children exhibited a reduction in the bias to choose the draw from the container with the smaller number of draws and the larger proportion of green to black lollipops compared to the same trial type in Experiment 1. These results suggest that rather than a seamless integration of both numerical and proportional information, the latter may exert a stronger influence on children's judgements. In order to investigate this possibility, I introduced a subtle change to the procedure of the next experiment: the containers remained covered until children made their choices. By doing this, I aimed to control for the potential over-salience of the large jars and their proportional information as opposed to the smaller, less-salient lollipop sticks in the cups.

Experiment 3

Methods

Participants

Data from 40 5- and 6-year-olds were included in the final analyses for this experiment (*Mean age*: 6;0, *range*: 5;0 – 6;9, *females:* 18). Children were recruited from and individually tested in local schools and museums in the Waterloo Region and the Greater Toronto Area.

Procedure and Design

The procedure was identical to Experiment 2 except for one change in the design. After the sampling process unfolds, instead of removing the grey boxes to reveal the distributions one final time as children made their choice between the 2-lollipop and the 5-lollipop draws, the containers remained covered until the end of each trial. The design was identical to Experiment 2 (see Figure 5).

<u>Results</u>

A GEE binary logistic regression with trial as a between-subject factor and age centred around the mean as a covariate revealed a significant main effect of trial on accuracy, *Wald* X^2(*df* = 3, *N* = 40) = 15.593, *p* = .001. No effect of age, nor interaction of trial by age, was found. Bonferroni-corrected pairwise comparisons revealed that performance on the 100:0 2v5 and the 88:12 2v5 trials were significantly different, *p* = .002. Figure 2.6 shows the proportion of correct responses on each trial.

On the 100:0 2v5 trial, where the outcome was guaranteed, all children selected the 2-lollipop draw. On the 50:50 2v5 trial, children responded correctly by choosing the 5-lollipop draw at a rate significantly greater than chance levels (*M*=92.5%, *SD*=.267, *p*<.0001). We also replicated the finding of the 88:12 2v5 trial in Experiment 2: children succeeded at choosing the 2-lollipop draw that has a higher probability of yielding a green lollipop in the (*M*=80%, *SD*=.405, *p*=.0002). Finally, on the 60:40 2v5 trial, children succeeded at choosing the 5-lollipop draw from the container with a lower proportion of green lollipops (M=85%, *SD*=.362, *p*<.0001), showing a significant improvement from performance on the same trial type in Experiments 1 (12/40 correct compared to 34/40 in Experiment 3, *p*<.001, Fisher's Exact test) and 2 (18/40 correct compared to 34/40 in Experiment 3, *p*<.001, Fisher's Exact test). While I should not make strong conclusions from these comparisons across experiments, this further supports the idea that keeping the containers visible as children made their choice may have made the proportional information overly salient and influenced their ability to integrate information on this trial type in Experiments 1 and 2.

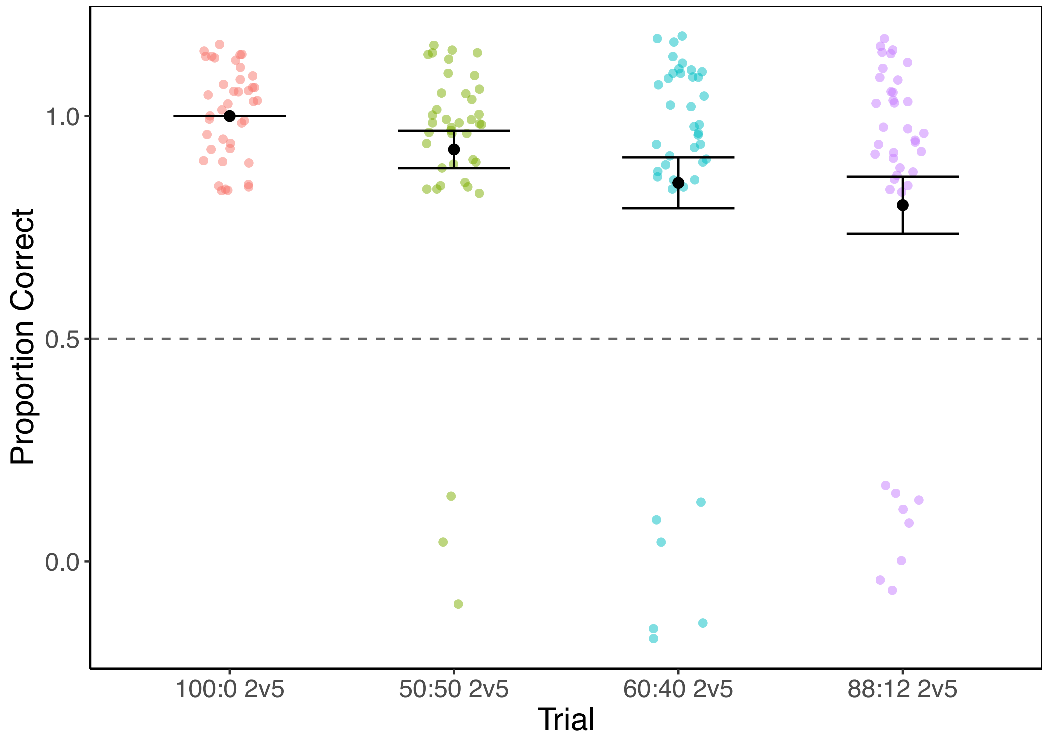

Figure 7. Mean correct responses on each trial indicated by dot (Experiment 3). Error bars indicate ± 1 standard error of the mean and individual data points are jittered for visibility. The dashed line represents chance (50%) performance.

Discussion

In this experiment, children's accuracy as a group was above chance on all trials. Of particular interest is the drastic improvement in performance on the computationally challenging 60:40 2v5 trial. In the previous experiment, children appeared to be choosing at random on that trial, whereas in this experiment, hiding the proportional information at the time of making a judgement seems to allow children to more effectively consider the numerical information in their decision. While this is speculative, it seems that there was some merit to the hypothesis that the over-salience of the distributions may affect the different weight assigned to proportional information in contrast with the numerical information.

In the final experiment with children, we will examine one more remaining open question in this line of work. Based on Experiments 1 – 3, it is now difficult to ascertain why children failed to correctly reason about the 60:40 1v3 trial in Experiment 1. They might have failed the trial because of a difficulty with the 1 vs 3 comparison, due to either task demands or a true inability to integrate the proportions and numbers. Alternatively, they might have failed the trial because the containers were visible and perhaps overwhelmed children's reasoning, biasing them too heavily towards the proportional information. These experiments are also missing a direct investigation of children's ability to make decisions based only on the proportional comparisons for the distributions in the 60:40 and 88:12 trials. In order to draw direct comparisons between children's choices when the proportions are held constant and the numbers of draws are manipulated, it is helpful to include trials in which the hand sampled 1 lollipop from each of the containers with the distributions from the 60:40 and 88:12 trials. Therefore, in Experiment 4 I will examine children's performance on the 88:12 1v3 and the 60:40 1v3 trials (as in Experiment 1) but with the proportions hidden (as in Experiment 3), and replace the 100:0 and 50:50 trials with an 88:12 1v1 trial and a 60:40 1v1 one.

Experiment 4

<u>Methods</u>

Participants

Data from 40 5- and 6-year-old children were included in the final analyses for this experiment (*Mean age*: 5;11, *range*: 5;0 – 6;11, *females:* 15). Children were tested in schools in the Waterloo Region.

Design and Procedure

The procedure was identical to Experiment 1 except for two changes in the design. First, just as in Experiment 3, the grey boxes that covered the proportions as the sample was drawn remained in place while participants made their choices, keeping the proportions covered until the end of each trial. Second, the 100:0 and 50:50 trials were replaced with trials in which single-items were drawn from the 60:40 and 88:12 proportions.

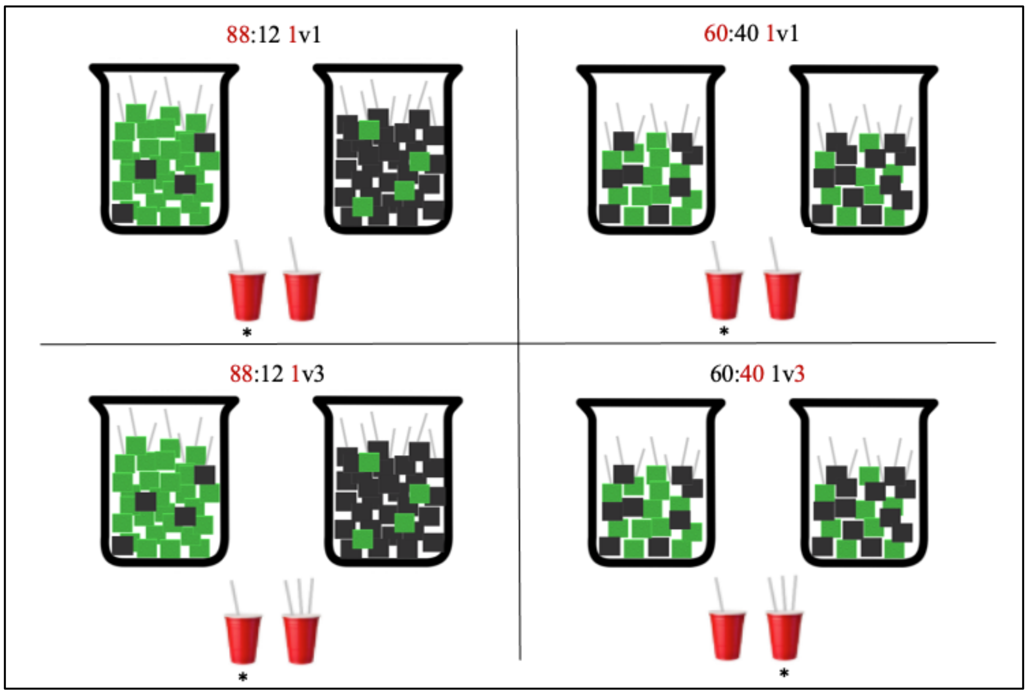

Figure 8. Depiction of the four trials in Experiment 4. Correct choices based on the computations in Chapter 1 are indicated by the asterisks.

Results

A GEE binary logistic regression with trial as a between-subject factor and age centred around the mean as a covariate revealed a significant main effect of trial on accuracy, *Wald* X^2(*df* = 3, N = 40) = 14.842, p = .002. Figure 2.8 shows the proportion of correct responses on each trial. No effect of age, nor interaction of trial by age, was found. Bonferroni-corrected pairwise comparisons indicated that the children performed better on the 60:40 1v1 than the 60:40 1v3 trial (p = .049). Also, performance on the 88:12 1v1 trial was significantly better than the 60:40 1v3 trial (p = .001).

A main interest is whether children performed above chance on each trial type. On the 88:12 1v1 trial, 92.5% of children selected the 1-lollipop draw from the majority green container (*Wald* $X^2(df = 1, N = 40) = 17.515$, p<.0001). On the 88:12 3v1 trial, 77.5% of children chose the 1-lollipop draw from the green majority container, which is correct in this case. (*Wald* $X^2(df = 1, N = 40) = 10.669$, p<.0001). On the 60:40 1v1, 82.5% of children chose the 1-lollipop draw from the green majority container, which is correct in this case (*Wald* $X^2(df = 1, N = 40) = 13.885$, p=.001). Finally, on the 60:40 1v3, 65% of children selected the 3-lollipop draw from the black majority container, which is correct in this case (*Wald* $X^2(df = 1, N = 40) = 3.487$, p=.062), not statistically different from chance.

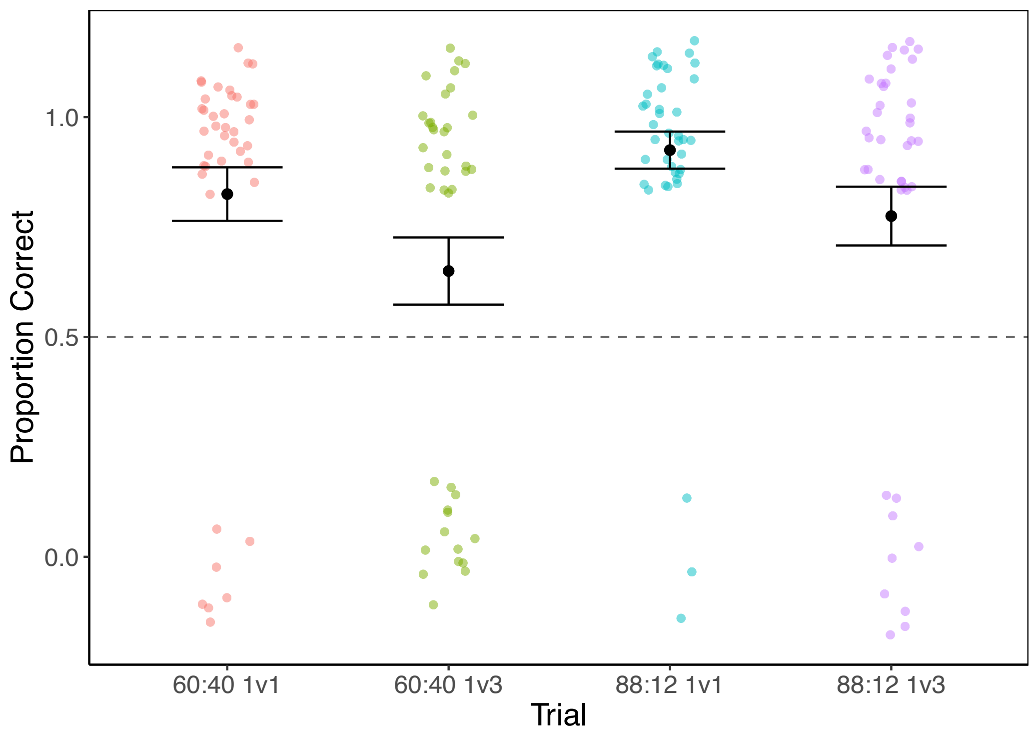

Figure 9. Mean correct responses on each trial indicated by dot (Experiment 4). Error bars indicate ± 1 standard error of the mean and individual data points are jittered for visibility. The dashed line represents chance (50%) performance.

Discussion

I conducted Experiment 4 in order to better understand the pattern of results found in Experiment 1. In particular, I wanted to ascertain whether children performed poorly on the 60:40 1v3 trial because of a difficulty with the 1 vs 3 comparison, due to either task demands or a true inability to integrate the draws and proportions for this trial type, or because the containers were visible and perhaps overwhelmed children's reasoning, biasing them too heavily towards the proportional information. Though this is speculative, and even though performance on this trial is still no different than chance, children showed an improvement on this trial in this experiment than the same trial in Experiment 1 where the proportions were visible at the time children made a selection (Experiment 1: 12/40 children chose correctly versus Experiment 4: 26/40; p=.003, Fisher's Exact test). Again, while I cannot make strong claims based on this cross-experimental comparison, it does appear that covering the distributions improves children's performance, despite their struggle with this trial in general.

Of interest is the difference in performance on the 60:40 trials, as well. Children's accuracy was significantly higher than chance on the 60:40 1v1 trial, when the samples were equal, but their responses did not differ from chance on the 60:40 1v3 trial. The difference in performance on these two trials was statistically significant, $p = .049$. By including the 1v1 comparison, it is easier to rule out the possibility that they could not tell the difference between the two distributions. This difference in performance is likely due to a failure of integration which intuitively makes sense as the extra piece of numerical comparison adds another layer of complexity in the decision-making process.

Experiment 5

Some of the trials we presented children in the previous experiments were quite challenging in terms of the computations involved and the relative probabilities that they have to consider. Would adults also be able to make some of the hardest comparisons? In this experiment, I tested whether adult participants will succeed at the tasks that I presented to children in the previous 4 experiments, in order to determine whether or not adults are capable of integrating information on the proportion of items and the number of draws in two competing location choices to make decisions that maximize their chances of getting targets. I also examined whether they do this at higher than chance (50%) levels on each individual type of trial and whether this varies across trial types that are objectively easier or harder to respond to.

<u>Methods</u>

Participants

Sixty-eight adults successfully completed this experiment (*M age:* 37 years*; females:* 22) on Amazon Mechanical Turk. Participants received $1USD for completing the experimental task, which took on average less than 5 minutes. An additional 12 participants were tested and their data was excluded due to failing the comprehension check (*n*=4); and failing at least one of the two predetermined exclusion trials (*n*=8). After piloting the experiment with 20 adults and having approximately 30% exclusion due to failing the comprehension check and/or exclusion trials, I set a target of recruiting 80 participants in order to end up with at least 50 participants.

Design and Procedure

Link to the task: https://uwaterloo.ca1.qualtrics.com/jfe/form/SV_7anZLiWjtkg5KSy.

Participants first read the instructions, which explained that they would play a series of guessing games and their goal was to get as many green lollipops as possible. Next they completed the

comprehension question used in Experiments 1-4, to ensure they understood that the goal. Any participants who responded to the comprehension question with anything other than "green" were excluded (n = 4). This was followed with a demonstration trial to allow them to practice making a selection (responses to this were not analyzed). They then saw 12 test trials in random order. On each trial, participants saw an image for 5 seconds with the two containers for that trial displayed. Then the containers were covered with grey squares and the items were sampled into the cups. Unlike in the experiments with children, any multi-item draws were done simultaneously, rather than one-at-a-time, to guard against boredom and disengagement. Adults made their selection by clicking on a cup. Adults saw every trial type presented to children in the previous four experiments, plus two pre-determined exclusion trials: 100:0 3v1 and 100:0 5v2. Both trials presented a larger number of items sampled from a container with all targets, and a smaller number of items sampled from a container with no target items. Participants who responded incorrectly to either one of them were excluded from final data analyses, as this indicated misunderstanding or ignoring the instructions, or random guessing. The experimental procedures and exclusion criteria were pre-registered here:
https://aspredicted.org/blind.php?x=ta9jk4

Results

A GEE binary logistic regression with trial as a between-subject factor revealed a significant main effect of trial on accuracy, *Wald* X^2(df = 10, N = 68) = 107.365, p < .0001. To follow up, I ran binomial tests against chance on every trial to determine whether participants performed significantly above 50% on each trial. These tests revealed that adults performed significantly above chance on all the trials except for the 60:40 2v5 trial, where only 36 of the 68 participants responded correctly (M=53%, SD=.503, p=.716).

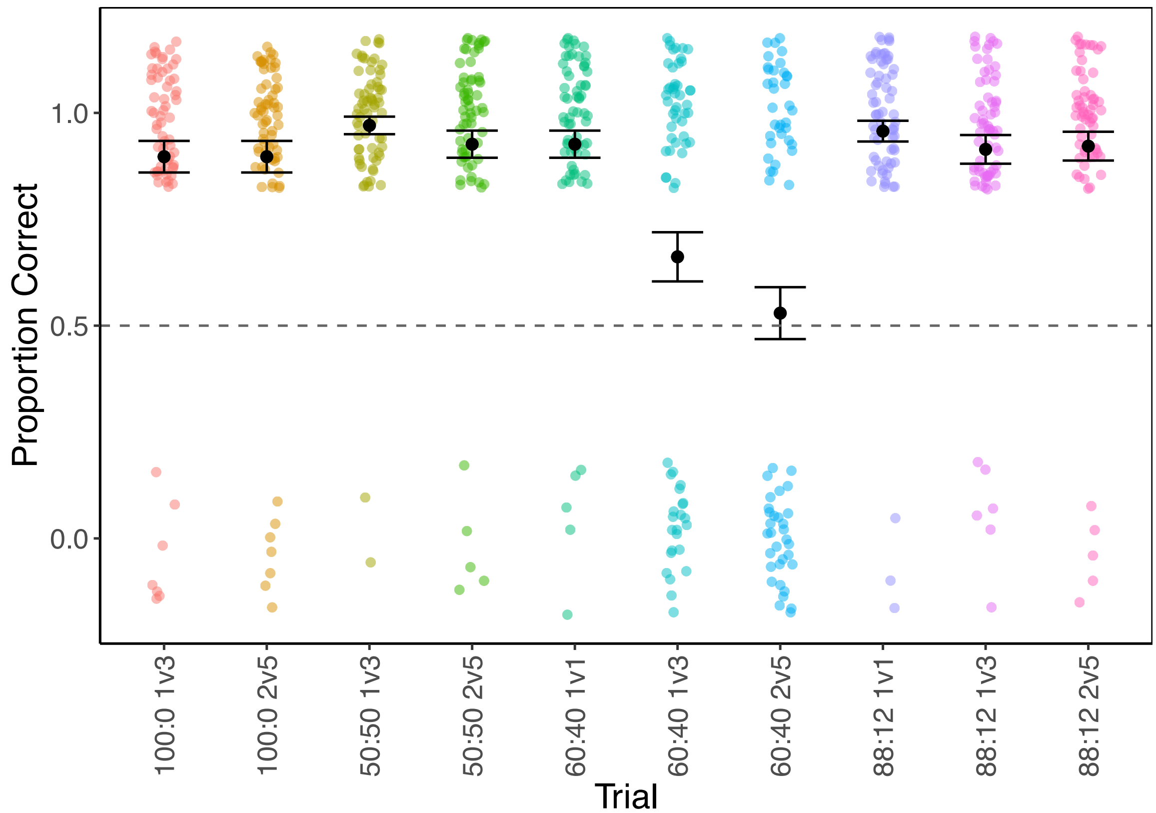

Figure 10. Mean correct responses on each trial indicated by dot (Experiment 5). Error bars indicate ± 1 standard error of the mean and individual data points are jittered for visibility. The dashed line represents chance (50%) performance.

<u>Discussion</u>

These results show that adults succeeded at all problem types involving integrating numerical and proportional information to maximize the probability of obtaining target objects, except for the 60:40 2v5 trial, where they responded no differently than chance.

Interestingly, data collected in the pilot, with an identical procedure (*N*=20) was promising in terms of performance on the 60:40 2v5 trial, with performance at 66%. One potential explanation for adults in the main experiment performing no differently than chance on this trial may be that the online testing using Amazon Mechanical Turk in this experiment introduced an additional source of noise to the data. This noise, when coupled with the difficulty of the 60:40 2v5 trial (computationally speaking), may have contributed to reduced accuracy on this trial in particular. Also in support of this interpretation is that adults responded correctly on the 100:0 1v3 and 100:0 2v5 trials only 90% of the time, whereas children never responded incorrectly to these trials, suggesting there is noise in the adult data. Another possibility is that the 60:40 2v5 trial is simply too difficult. Although it then becomes hard to know why children performed above chance on it in Experiment 3. This adult experiment did provide some clarity on the second-most challenging trial, 60:40 1v3. Even adults tested online responded at above chance-levels as a group, albeit this ability was not particularly strong, suggesting that while this trial is difficult, people can determine the better choice. It also raises a question regarding statistical power in my child experiments; namely, the 66% of adults responded correctly to this trial (statistically different from chance with 68 participants), whereas 65% of children responded correctly on this trial (not statistically different from chance with 40 children). This highlights the possibility that had more children been tested, their performance may have differed from chance.

A number of follow-ups to this adult experiment can be conducted in order to better understand these findings and their implications. Firstly, it is likely that the online setting of this experiment created a significantly different testing experience to the in-person alternative, which was designed and had ethics clearance and then was discontinued when the COVID shutdown occurred (3 participants had been tested). Secondly, the decrement in performance on the 60:40 2v5 trial in this experiment, a trial that children succeeded on in Experiment 3, could be due to the benefit of sequential (“experienced”) sampling as opposed to simultaneous (“summary”) sampling, in probability experiments. This could be due to a phenomenon known as the "description-experience" gap, which has been demonstrated in adults a number of times, and can be further investigated in future work (for a review see Schulze & Hertwig, 2021).

General Discussion

Across five experiments, I established that 5- and 6-year-old children and adults can – under certain conditions – integrate two sources of numerical information to make probabilistic decisions. Across all the experiments, performance on trials where the proportion of target to non-target objects is deterministic (100:0), performance is close to, or at, ceiling levels. Performance on the 88:12 trials, where choosing the smaller number of draws from the container with mostly target objects would maximize their chances of getting green, was also significantly better than chance levels across all the experiments. When presented with equal proportions of black to green lollipops, children appeared to be implementing a proportions-focused strategy when proportions were visible during choice in the 1v3 comparison (Experiment 1) but appeared to correctly choose the larger number of draws when proportions were hidden in all subsequent experiments, suggesting that hiding proportions during choice improved performance. The 60:40 1v3 trial appeared to be the most challenging, both for children and adults. Most children chose the 1-lollipop draw from the green-majority container on this trial in Experiment 1 and were no different than chance on the same trial in Experiment 4. Choosing the 3-lollipop draw from the black-majority container was also challenging for adults, though they did so at above chance levels. However, choosing the 5-lollipop draw, as compared to 2 proved especially difficult for adults, as this trial was the only trial where adults' accuracy was no different from chance. Curiously, children succeeded in choosing the 5-lollipop draw at above chance levels on this trial type when the proportions were hidden at test.

This set of experiments highlights children's sensitivity to perceptual salience as evidenced by performance on Experiment 3 compared to Experiment 2. While this is largely speculative, it appears performance on the 60:40 1v3 trial, which is very mathematically challenging, improves from Experiment 1 to Experiment 4, as well. This may be because in Experiment 1, where the distributions were visible, children found it difficult to ignore the greater proportion of target objects and opt instead for the larger sample from the distribution with a lower proportion of target objects. Interestingly, when comparing across experiments, children's accuracy on that trial went from 30% on Experiment 1 to 65% on Experiment 4 when I simply hid the proportions by covering them with an opaque box at the time where we prompted children to make a decision. Their choices seemed to be sensitive to this minor change in design. I would not make strong claims from comparing across experiments, but a suitable follow-up to this speculation would be to manipulate whether the proportions are visible (or hidden) at the time of making a choice in a within-subjects design. This would reveal whether or not performance on the same trial type would be significantly different with proportions being visible, or hidden.

Another natural follow-up to this work would be to manipulate proportions and distributions more systematically. The trials always involved one of four distributions: 100:0, 88:12, 60:40, and 50:50. Adding more layers to this would be helpful in better understanding the extent of children's (and adults') sensitivity to these differences. Closely related to that, the computed probabilities (taking into account both proportional information as well as sample number) can also be systematically manipulated such that there is a larger distribution of competing probabilities. These possibilities will be discussed in further detail in Chapter 5.

Some trials are simply more (or less) challenging based on the probabilities and this should also be investigated further. For instance, one particularly important feature of the characteristic of the different trials is whether the computed probabilities of getting the target object between the competing samples lie on the same or opposite sides of 50%. When a comparison lies on the same side of 50%, this is notoriously difficult for children to deal with (Falk et al., 2012). In the 88:12 1v3 trial, the competing probabilities of getting at least one green lollipop are 88% versus 34%. On the other hand, the 60:40 1v3 trial involves the comparison between 60% and 81%. For starters, the quantitative difference between these competing probabilities (i.e., the relative likelihood or "ratio of ratios") contributes to the difference in difficulty of these two trials (i.e., 88% and 34% are further apart than 60% and 81%). But also, when both probabilities lie on the same side of 50%, one cannot adopt a very useful heuristic to make a choice, which is to determine whether an individual sample container is "good" or "bad" and choose based on this. When comparing 60% and 81% both of these are good, or winners, because they are both more likely than not to produce targets. In contrast, when one sample container is a clear "loser" because it is most likely to not contain targets, it is easy to avoid this sample and choose the winner. I will return to this point in particular to discuss in more depth in Chapter 5, as this is relevant to all chapters of the dissertation.

Research on children's abilities to integrate multiple pieces of information in decision making extends to non-numerical domains as well (Bridgers et al., 2016; Gualtieri et al., 2020; Sobel et al., 2004). When presented with two sources of information, where one is numerical in nature and the other is social (e.g. testimony), children show evidence of considering both sources of information when making judgements: Children as young as 4 can integrate observed causal frequency information with an informant's competing testimony, by weighing the strength of each piece of information (Bridgers et al., 2016). That is, if a child must decide which of two objects is a blicket (an item that activates a detector), they integrate a person's testimony with their own observations of the causal efficacy, relying more on a confident person than a non-confident person when that person's testimony contradicts the data they have observed. There is also evidence that starting at the age of 4, children use prior knowledge to inform the hypothesis space when making similar inferences under uncertainty (Sobel et al., 2004). In particular, children take into consideration the base-rate proportions of target options in their judgement of whether or not an object is a blicket). When an object displayed unambiguous causal function, all children judged it to be a blicket, regardless of whether they were told that blickets were rare or common. However, when another object was ambiguous in whether or not it was causally functional (i.e., the activation could have been caused by that object, but another object was also present that could have caused the activation), 4-year-olds were more likely to determine that they were blickets if blickets were common than if they were rare. Thus, the experiments in this chapter add to evidence of integration that fits with a broader line of research suggesting that young children can integrate multiple sources of numerical information when making inferences under uncertainty.

In sum, the results from chapter 2 reveal that children and adults can engage in some integrative reasoning that allows for the consideration of numerical and proportional information when making probabilistic decisions. This makes sense considering that young children, as well as adults, use complex statistical information to understand their environments and navigate the world around them (Aslin et al., 1998; Saffran et al., 1996). Even infants have shown a sensitivity to proportional and numeric information (Denison et al., 2014; Feigenson et al., 2004; Kayhan et al., 2017; Odic, 2017; Teglas et al., 2007; 2011; vanMarle & Wynn, 2011; Xu &

Garcia, 2008). This, in addition to the success of children in Experiments 1-4, was the primary motivation for taking a leap backwards in development and investigating these abilities in infants in the next chapter.

Chapter Three: Integration of Number and Proportion in Infants

Beginning in infancy and developing throughout childhood, children have remarkable abilities to reason about numbers and proportions (e.g., Feigenson et al., 2004; Gallistel & Gelman, 2005; Gelman & Gallistel, 1978; Dehaene, 1997, 2009; Piaget, 1965; Xu, 2003). Children, like adults, are capable of estimating large numerosities (e.g., Izard et al., 2009; Lipton & Spelke, 2003; Xu, 2003; Xu & Spelke, 2000; Xu et al., 2005), as well as keep track of and compare smaller collections of individual items (e.g., Feigenson et al., 2002; Starkey & Cooper, 1980). In addition to being able to represent small and large numbers, infants' numerical representation systems even allow them to engage in seemingly complex numerical comparisons and computations (Kobayashi et al., 2004; 2005; Koechlin et al., 1997; McCrink & Wynn, 2004, 2000). In this chapter, I examine whether infants can combine numerical information (in the small number range) and proportional information together in a probabilistic decision-making task.

In terms of small numbers, even infants as young as 10 months can precisely represent, track, and explicitly choose between sets containing three or fewer items (Antell & Keating, 1983; Feigenson et al., 2002; Starkey & Cooper, 1980; Starkey et al., 1990). In looking time studies, when infants were habituated to one set size (e.g., 2) and then shown a novel set size (e.g., 3), they looked longer at the novel set size, suggesting that they could discriminate between small numbers of objects (Starkey et al., 1990). Additionally, a series of now classic experiments examined whether or not infants can track and then correctly choose the larger of two quantities of hidden objects. Infants were shown two quantities of crackers that were individually placed in separate opaque containers, and reliably crawled to the container with the larger number of crackers when the number of crackers being compared was 3 or fewer (Feigenson et al., 2002;

vanMarle, 2013). We also know infants are able to create working memory models of at least two sets of 3 or fewer items, and can compare these models on the basis of one-to-one correspondence to determine numerical equality or ordinal relations (Le Corre & Carey, 2007; Feigenson & Carey, 2003; Feigenson & Carey, 2005; Feigenson & Halberda, 2004).

Infants are capable of discriminating between sets on the basis of number as well as other non-numeric properties of the stimuli. For example, infants as young as 7-months-old respond to changes in various quantitative dimensions in visual displays, such as contour length (i.e. the sum of the perimeters of the individual objects in the set; Clearfield & Mix, 1999), with items up to a limit of 3. That is, after habituating to either 2 or 3 squares of the same size, infants were then shown both the same number with a new contour and the same contour with the old number. By comparing the sum of the continuous properties that were bound to the visual stimuli, they were able to detect the change in contour length even when number was held constant. Similarly, by comparing the models for one-to-one correspondence, infants can also detect that two sets no longer match in numerosity and dishabituate to the test trial that differ on the basis of number when contour length is matched (Cordes & Brannon, 2009a; Margolis, 2020). This shows that they demonstrate this ability on the basis of number *as well as* other continuous variables.

Infants, children and adults also have a basic ability to represent large numerosities: the approximate number system (ANS). This system produces rapid and inexact representations of numerosities that are too large to count (and in the case of infants and non-human animals, in the absence of the ability to count). These representations adhere to Weber's Law, such that the ability to discriminate between two large sets of objects (based on number) depends on the ratio of their magnitudes (Halberda & Feigenson, 2008; Lipton & Spelke, 2003; 2004; Whalen et al., 1999; Xu & Spelke, 2000). This ability to discriminate large numbers is generally tested by comparing two sets or arrays of objects, and the acuity with which these judgements are carried out does not depend on the total number of objects in the array, but rather on the ratio between the objects presented (i.e. the closer together the two numbers are, the more difficult it is to discriminate them from one another). This acuity steadily becomes more fine-tuned throughout childhood and across development (Odic et al., 2013). While adults discriminate numerosities at about a 1.15 ratio (Van Oeffelen & Vos, 1982), 6-month-old infants discriminate numerosities that differ by a factor of 2.0 (e.g. 4 vs. 8, but not 4 vs. 6), and 9-month-old infants discriminate numerosities that differ by a factor of 1.5 (e.g. 8. vs. 12, but not 8 vs. 10) (Lipton & Spelke, 2003, 2004; Xu, 2003; Xu & Spelke, 2000).

There is a surprisingly small amount of work looking at ratio representation and discrimination, or proportional reasoning with large numbers of items, in infants and young children. From the little work that does address this question, it appears that this idea of discrimination acuity that gradually improves over development extends to reasoning about the ratio of ratios, as well. McCrink and Wynn (2007) found that 6-month-olds were able to discriminate between ratios that differed by a factor of 2 but not 1.5. This suggests that infants are capable of ratio abstraction, whereby they can be habituated to a particular ratio by viewing displays that differ in overall numbers of objects but have constant ratios and they will dishabituate to ratios that are sufficiently different (by at least 2 times). It also appears that infants may be capable of using that representation of the ratio of a distribution when determining the match between a population and sample. One of the only studies that has examined whether there is development in infants' relative likelihood estimation abilities comes from a probability study with 6- to 18-month-old infants where multiple pairs of samples were drawn from a distribution such that the magnitude of difference in likelihoods between each pair of samples varied. Kayhan et al. (2018) investigated whether infants are sensitive to the difference between likelihoods of multiple outcomes by trying to determine if infants would distinguish between a pair of outcomes that differed only slightly in probability of occurring (statistically-speaking) and another pair of outcomes where the difference in relative probabilities was much larger. Results showed that as infants aged, their ability to make probability judgments across different relative likelihoods became more fine-grained, though the results from this study were not entirely straightforward. (Kayhan et al., 2018).

Despite the lack of research specifically examining ratio abstraction in infancy, there is a line of research that suggests infants (and non-human animals) can use ratio information to make probabilistic inferences. When 8-month-olds are presented with a container of mostly red balls and a few white balls (70 red:5 white, for example), they looked longer when a small sample drawn randomly from the box did not match this distribution (i.e. contained more white balls than red balls) than when the sample matched the distribution (i.e. contained more red balls than white ones; Xu & Garcia, 2008). This suggests that they inferred the probability of the sample given the population and when this sample did not reflect what they observed in the population, this violated their expectations. Similarly, when 12-month-olds saw three yellow objects and one blue object randomly shuffled inside a container before one of them exited, they looked longer when they observed an unlikely event (i.e., a blue object leaving the container) than when they observed a likely event (i.e., a yellow ball exiting the container; Teglas et al., 2007). Converging evidence from various labs suggests that infants have expectations about both these multi-item sampling events and single sampling events, at least as measured by looking-time tasks (Denison et al., 2013; Kayhan et al., 2018; Lawson & Rakison, 2013; Teglas et al., 2007, 2011; Xu & Denison, 2009; Xu & Garcia, 2008).

These looking-based experiments established that infants can observe a population and then reason about whether a sample was probable or improbable. A similar line of research examines whether infants and non-human animals can make predictions about sampling events based on the distribution of items in a population, and when they must decide which of two competing populations is most likely to yield a reward. For example, they are able to predict which of two random draws would be more likely to yield a particular object based on the proportion of objects in the populations from which they were drawn (Denison & Xu, 2010, 2014). This work demonstrated that infants are not only able to make predictions based on proportional information, but are also able to use these predictions to guide their actions.

In single-draw probabilistic reasoning tasks, infants and apes appear to have sophisticated abilities to make accurate inferences. When infants were shown two transparent containers that have target and non-target objects in different ratios, and the experimenter sampled from these containers by placing a lollipop from each jar in an opaque cup without allowing the infant to see its color, infants reliably walked or crawled towards the sample that had come from the jar with the higher ratio of their preferred lollipop (Denison & Xu, 2014; see Rakoczy et al., 2014 for very similar evidence from non-human primates). This persisted even when the jar with the higher ratio of target to non-target lollipops contained a smaller absolute number of target lollipops. In addition, infants have also shown that they can integrate a number of different factors, quantitative and otherwise, when determining the relationship between a sample and a population. For example, 11-month-olds are capable of weighing an experimenter's preference for a particular type of object with the proportion of that object to other ones to infer the outcome of a draw (Xu & Denison, 2009). If the experimenter establishes a preference for red balls, for example, and draws mostly red balls from a distribution with few red balls and mostly white balls, infants are not as surprised by this as they are when the same sample is drawn by a blindfolded experimenter. However, there is no research on infants' ability to integrate two numerical dimensions, such as numbers and proportions, to make probabilistic decisions.

Given that infants appear to be able to reason about both numbers and proportions independently of one another, I aimed to examine if they can integrate this information. While this is a challenging task, the results obtained from school-aged children and adults in Chapter 2 showed some promise with regard to performance on tasks that require this kind of flexibility in using both numerical as well as proportional information to arrive at a favourable outcome. In order to examine these integrative abilities further and, in particular, investigate their developmental origins, I used a similar paradigm to the one used in Chapter 1, modified for use with 10-12-month-old infants. There were two important considerations in designing a physical version of the lollipop game to use with infants. First, I kept distributions hidden when asking infants to choose between samples. This design seemed to work better with older children and most infant and ape paradigms in the probability literature use this approach (e.g., Denison & Xu, 2014; Rakoczy et al., 2014). Second, we only presented infants with the trials where the number of lollipops sampled were either 1 or 3. The reason for doing this was two-fold: first, because this was a non-verbal paradigm, the possibility of the instructions biasing or confusing the participants in some way was eliminated, so there is no concern that infants might think the experimenter is suggesting they choose the single or multi-item container based on her instructions. Second, this avoids crossing the boundary between small and large numbers established in previous research with infants this age in terms of the comparison between the number of draws in the sample (2 versus 5 would cross this boundary, whereas 1 versus 3 does not; Feigenson & Carey, 2005; Feigenson et al., 2002). In the following experiments, I sought to assess the probabilistic decision making using numerical and proportional information is infants.

Methods

Participants

Data from 46 10-12-month-old infants were included in the final analyses for this experiment (*Mean age*: 11;15 [months; days], *range*: 10;9 – 12;29, *females:* 25). An additional 22 infants were tested and their data were excluded for failure to select the pink lollipop during both preference checks (*n*=15); inability to crawl to one of the cups on one or more trials (*n*=5); and experimenter error (*n*=2). Children were tested in the Developmental Learning Lab at the University of Waterloo.

Procedure

An experimenter first established that there is a favourable target object (in this case, a pink, sparkly, light-up lollipop) and an unfavourable one (a black lollipop). This was done using verbal cues such as "I like this one," referring to the pink lollipop, and non-verbal cues, such as smiling when presenting the pink lollipop to the infant. In contrast, the experimenter frowned and said "I don't like this one" when referring to the black lollipop. In previous work, infants have generally preferred the pink lollipop at baseline, with no cues from the experimenter. In this experiment, I wanted to be as certain as possible that infants had a goal for the pink target, so I added these cues and only included infants who clearly preferred pink. Following this, there was an initial preference check where the experimenter placed both lollipops on the floor to ensure that all infants crawled towards the favourable object (the pink lollipop, in this case). Infants' choice was recorded based on which lollipop they crawled to.

Following this, infants were presented with two trials in counterbalanced order: 88:12 1v3 and 60:40 1v3. Each trial started with the two containers covered. The experimenter then placed two empty cups in front of the containers. The covers were removed and the containers were then displayed one at a time to the participant so that they could see the distributions. The containers were then covered and the experimenter drew either 1 or 3 lollipops from each container and placed them in the cups, such that the participant was only able to view the sticks in the cups, not the colours of the lollipops. In the case of the 3-lollipop sample, the experimenter sequentially sampled one lollipop at a time. The experimenter then prompted the infant to crawl towards the cup of the lollipops. Afterwards, infants were presented with a second preference trial in which the experimenter again presented both the black and pink lollipops, placed them on the floor, and prompted the infant to crawl to one of them.

Design

Each infant was presented with a preference trial, followed by two test trials presented in counterbalanced order, then a final preference trial at the end. If an infant crawled to the black lollipop on the first preference trial, it was repeated 2 more times (for a maximum of 3 initial preference trials). Any infants who failed to crawl to the pink lollipop at least once on the initial preference check trials *or* on the final preference check were excluded from the final sample. While we could have implemented a more stringent inclusion criterion, we did this primarily to ensure that we did not have to exclude a large number of infants, who are typically difficult to find and recruit. Please see Figure 11 for a schematic of the experimental design, and Figure 12 for the test trial types included.

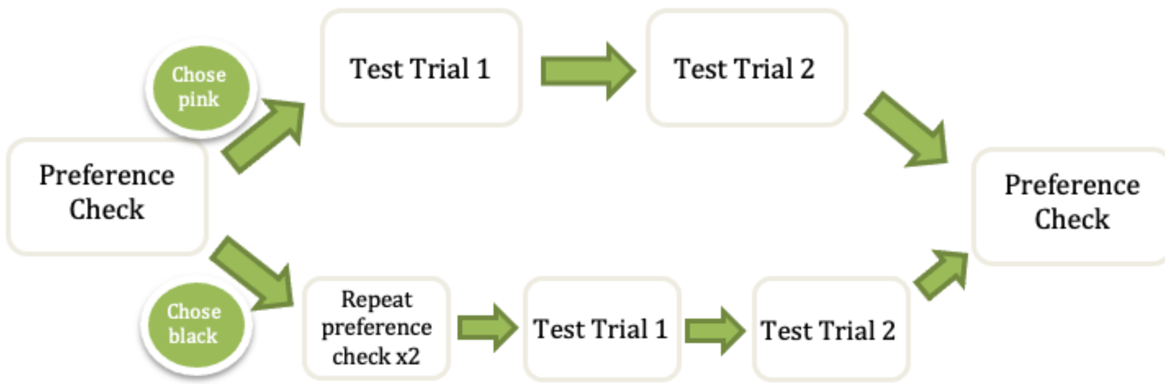

Figure 11. *Schematic of the experimental design*

Figure 12. Trials presented to infant participants (order of presentation was counterbalanced across participants). Proportion of target to non-target objects is indicated below each container. Correct choices based on the computations in Chapter 1 are indicated by the asterisks.

<u>Results</u>

A GEE binary logistic regression with trial as a between-subject factor and age centred around the mean as a covariate revealed no significant main effect of trial on accuracy, *Wald* $X^2(df = 1, N = 46) = 1.37, p = .242$. There was no effect of age or an age by trial type interaction. This suggests that infants did not reliably use either or both cues to make selections, as their responses were very close to chance-based selection.

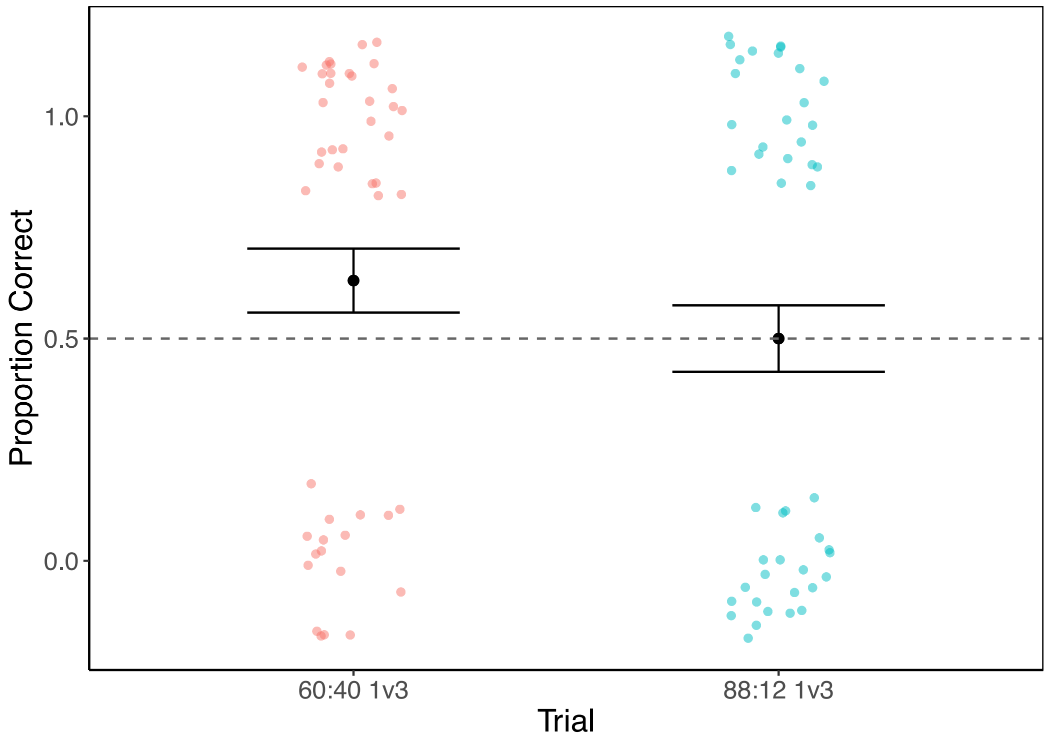

Figure 13. Mean correct responses on each trial indicated by dot. Error bars indicate ± 1 standard error of the mean and individual data points are jittered for visibility. The dashed line represents chance (50%) performance.

<u>Discussion</u>

Results showed that infants were not able to use the numerical and proportional information to make probability judgments. This pattern of results could be caused by a number of factors. First, it is possible that there is development in these abilities across the age group, which the experiment did not have the power to detect and reveal. A second possibility is that these infants (or at least, a subset of them) had a different goal in mind than the one we intended. Finally, it is possible that the task is simply too difficult for the age group included in this experiment (10- to 12-month-olds).

To address the first possibility, it may be the case that the number of infants tested was simply too small to reveal any age effects. We tested 46 infants with roughly equal numbers of 10-, 11-, and 12-month-olds (approximately 15 each). If there are indeed age effects in this range, it is possible that we did not detect them because of the small number of infants tested. Previous work has shown that there is development during this age in the probability paradigm used by Denison & Xu (2014). In that paper, only the oldest half of infants (those over 11 months) were able to succeed at the most challenging case in the final experiment, where infants had to make probability comparisons on the same side of 50% (see the General Discussion of Chapter 2 for why this poses a particular challenge). If infants are still developing their probabilistic reasoning abilities during this age range, it is not surprising that they may struggle to integrate proportions with number draws at this stage.

Another important issue to raise is that while I was careful to not require that infants compare individual draw numbers that crossed the small-large boundary, this task may require them to integrate a (perhaps ANS-based) large number ratio representation with a small number (the draws) to make the individual computations for each container and its respective sample. This is the first investigation, to my knowledge, of such a skill. Given that infants have difficulty comparing sets of items that cross between the small and large number ranges (<3 compared to >3; Feigenson et al., 2002; vanMarle, 2013), it is possible that the system representing the ratios (perhaps the ANS) and the system representing the number of draws (parallel individuation) cannot be combined in such a way. Therefore, perhaps we should not have even expected infants to be able to integrate an abstracted ratio with a small number. I will return to this point more fully in Chapter 5.

Based on past research of numerical reasoning and probabilistic inferences in infancy, we recruited participants in what may have been the earliest possible age to observe these abilities, admittedly however, this was probably ambitious. Additionally, it is also not outside the realm of possibility that, had we also tested more infants across the entire age range, I could have found some kind of age effect, either a main effect of age or an interaction between age and trial type. As mentioned earlier, there appears to be development between 10-12 months based on previous work (Denison & Xu, 2014). It is possible that infants at different ages could be employing different strategies or that individual infants might be using different strategies that in turn wash each other out. For example, some individual infants might focus on proportions, while others may rely more on number. We do know that infants are able to perform complex computations with both small and large numrosities, though. They can, for example, engage in non-symbolic approximate addition and subtraction on large-number arrays (McCrink & Wynn, 2004). However, it is possible (if not likely) that the ability to readily integrate multiple different and complex pieces of numerical information to make judgements about relative probabilities develops later.

Finally, a number of methodological concerns need to be addressed. While we were able to verbally instruct the school-aged children to try and maximize their chances of getting green, and framing it such that the game-like scenario provided sufficient motivation to do just that, the same could (for obvious reasons) not be done with the infants in this experiment. We hoped that the verbal and non-verbal cues that the experimenter provided the infants were sufficient to bias them to prefer the pink lollipop and recognize that obtaining it was the goal, however this may not have been the case. If this is true, performance may then not be entirely reflective of infants' integrative abilities (or lack thereof), but it could rather be a reflection of different infants being motivated to seek out a goal other than maximizing the chances of getting pink lollipops (e.g., for some infants, just getting more lollipops of any colour might have been very rewarding).

The results observed in this chapter indicate that more groundwork needs to be done in the study of integrating information on small numbers and proportions in making probabilistic inferences. There are a number of potential follow-ups that can help in better understanding the mechanisms at work in tasks like this one. One possibility would be to systematically test different proportions and draw numbers, as suggested in chapter 2. Another interesting follow-up that would be informative with this age group especially would be to employ an eye-tracking methodology. This would be an interesting avenue to explore, in particular because it would be informative in understanding infants' allocation of attention in this task. Would they look first at the proportions of the distributions to abstract the ratio information, then at the sampled objects before making a choice? Or would there be a pattern of back-and-forth oscillation between fixating on the proportional information and the sample numbers? Do they consider each sample-proportion pair individually and compare them to one another, or do they compare proportions first before moving on to compare sample numbers? And finally, would infants make accurate anticipatory looks when predicting where an agent might go to collect target items? Like other tasks, it is possible that infants would show competence earlier in a looking-based experiment than in one that requires they plan and execute a motor act.

This experiment highlighted a missing step between the relatively robust findings of early school-aged children's ability to integrate numerical and proportional information in probabilistic decision making and the inability to demonstrate this ability in infancy. This big leap backwards from school-age children to infants was prompted by the success of older children on these tasks and the literature suggesting that infants may have the abilities to do each aspect separately (the number comparison of the draws and the proportion comparison in the containers). In the next chapter I turn to an intermediate age to determine if the abilities shown by 5- and 6-year-old children (and adults) in Chapter 2 extend to a younger age group of 18-30-month-olds.

Chapter Four: Integration of Number and Proportion in Toddlers

The results from the infant experiment revealed that while infants may have the requisite abilities to use both numerical and proportional information to make decisions, integrating number and proportion in probabilistic decision making appears to be beyond their capabilities. On the other hand, the data from school-aged children suggests that 5- and 6-year-olds are able to engage in this kind of flexible data integration in order to make decisions and maximize their chances of arriving at a favourable outcome, at least under some conditions. One possibility is that these abilities emerge right around the age of our youngest participants in the experiments with school-age children (around the age of 5), however, it would be interesting to explore another possibility, which is whether these abilities emerge sometime between the infant and school-aged years. Much like with infants, there is no literature examining this integration in the toddler years.

In fact, the literature on the development of numerical and proportional reasoning in toddlerhood is quite sparse. Generally speaking, toddlerhood (from roughly 18 -30 months) is relatively understudied compared to the earlier infancy period (0-18 months) and preschool ages (3-6 years) in much cognitive development literature. In terms of probability, specifically, a review of the literature suggests that there are almost no studies in this age group. One exception is work by Waismeyer et al. (2015) examining causality with 2-year-olds by presenting them with a probabilistic causal system. This is quite different from the sampling studies that are most relevant to this dissertation; however, it does provide evidence that 2-year-olds can consider two likelihoods simultaneously to make an inference about which of two items is more likely to activate a machine. There are a number of potential reasons for this gap in the literature. One major reason is that it is difficult to design experiments suitable for this age range. While older children may easily understand and respond well to explicit instructions in a game-like context, for example, this may not be true for toddlers who are less skilled in their language comprehension and who may have less experience with structured, instructions-based activities. On the other hand, using nonverbal experimental designs commonly-used in infant paradigms may not be engaging enough to elicit participant responsiveness and attentiveness in toddlers, at least in the context of studying numerical processing.

In terms of the numerical reasoning literature, there is some work suggesting that children starting at 24 months of age have mastered representations of small numbers and can represent ordinal relations between them (up to a limit of 5), even when controlling for continuous variables such as surface area (Brannon & Van de Walle, 2001; Huntley-Fenner & Cannon, 2000; Sophian & Adams, 1987; Strauss & Curtis, 1984). In other words, toddlers understand that a set of 5 objects is different and greater than a set of 2 objects, for example. This sheds some light on toddlers' ability to make numerical comparisons, in particular ordinal judgements, before mastering the verbal counting system that some believe may mediate these judgements in older children (Brannon, 2002). Even younger toddlers, 14-16-month-olds, appear to engage in reasoning about ordinal relationships between numerical values in implicit, looking-based tasks. After habituating to pairs of stimuli that differ in absolute numerosities but had the same constant ordinal relationship, (e.g. 1 & 2, 3 & 4), toddlers dishabituated when they were shown test trials with a reversed ordinal relationship, or no ordinal relationship (i.e. equal numerosities), but not if they were shown test trials with different numerosities while keeping the ordinal relationship consistent (Cooper, 1984). Although this is a related, not identical concept, this work suggests that children in this age range have a solid ability to compare what is "larger" and "smaller" and the relations between sets of numbers, at least up to a limit of 5 objects. In terms of large number comparisons, children as young as 2.5 years of age are able to discriminate between large numbers that differ by ratios between 2:3 and 1:2 (Halberda & Feigenson, 2008; Rousselle & Noël, 2008). Children at this age perform more poorly on comparisons of 15 vs 18, for example, than on comparisons of 6 vs. 9, suggesting that their accuracy on these comparisons was controlled by the ratio of the numerosities and not by the absolute difference between them (Cheung & LeCorre, 2018).

In sum, little work has so far examined probabilistic reasoning in the toddler years, while there is slightly more work on basic numerical abilities. This, as well as the success of school-aged children and failure of infants in my tasks, prompted my decision to move on to testing toddlers in this chapter using a very similar paradigm to the one used with infants, continuing the work that was started with the school-aged children's studies. If there is development in the months immediately following the 12-month mark in the domains of numerical comparisons and ordinal knowledge, or even in simple probabilistic reasoning as suggested by the age effects found by Denison & Xu (2014) then there may be reason to believe that testing toddlers may help bridge that gap between the numerical/proportional integration observed in older children and adults, and the lack thereof in infants under 12 months.

Methods

Participants

Data from 40 toddlers (*mean age* = 20;30 [months; days], *range*=18;6-27;4, *females*: 19). An additional 17 toddlers were tested and their data were excluded for fussiness and failure to complete the experiment (*n*=8); failure to select the pink lollipop during both preference checks (*n*=6); experimenter error (*n*=2); and parent interference (*n*=1). Children were tested in the Developmental Learning Lab at the University of Waterloo.

Procedure

I used the same physical game task used with the infants in the previous chapter with the slight modification of asking participants to point to their choice rather than crawl to it, which seemed like a more appropriate set-up for this age group. Participants were tested in a choice paradigm in which they sat across a table from an experimenter with their parent.

I presented the toddlers with the same trials that were used with the infants in the previous chapter, using the same experimental stimuli and script. I first established that the pink, sparkly lollipop is the favourable object by using verbal cues such as "I like this one," and non-verbal cues, such as smiling when presenting the pink lollipop to the toddler. I also established that the black lollipop is unfavourable by frowning and saying "I don't like this one" when referring to it. Following this, there was an initial preference check where the experimenter placed both lollipops on the table within reach of the toddler and asked "which lollipop do you want?" Toddlers' choice was recorded based on which lollipop they touched or pointed to first. Following this, toddlers were presented with two trial types in counterbalanced order. These trials were identical to the ones used in the infant experiment. Each trial started with the two containers covered on the table. I placed two empty cups in front of the containers. I removed the covers and the containers were then displayed one at a time to the participant so that they can see the distributions. I then covered the containers and drew either 1 or 3 lollipops from each container and placed them in the cups, such that the participant was only able to view the sticks in the cups, not the colours of the lollipops. I then prompted the toddler to choose one of the cups by moving them both within reach of the toddler and asking them which cup they wanted. After both test trials, I presented toddlers with a second preference trial during which I presented both the black and pink lollipops, placed them on the table within reach of the toddler, and prompted the child to choose one of them.

Design

The design was identical to the experimental design used in the infant experiment. Participants were presented with a preference trial where the experimenter used the same verbal and nonverbal cues previously used with the infants to bias the child to choosing the pink lollipop. We then presented them with the two test trials in counterbalanced order, followed by a final preference check.

Results

A GEE binary logistic regression with trial as a between-subject factor and age in months centred around the mean as a covariate revealed no main effect of trial on accuracy, *Wald* $X^2(df = 1, N = 40) = 3.487$, $p = .062$. No effect of age was found, nor was there a trial by age interaction.

Binomial tests for each trial type showed that toddlers succeeded at choosing the draw that yields a higher probability of obtaining the pink lollipop in the 60:40 1v3 trial, (M=82.5%, SD=.385, $p < .0001$). They did not, however, choose the draw that yields a higher probability of obtaining the pink lollipop in the 88:12 1v3 trial, at above chance levels (M=62.5%, SD=.490, p=.154).

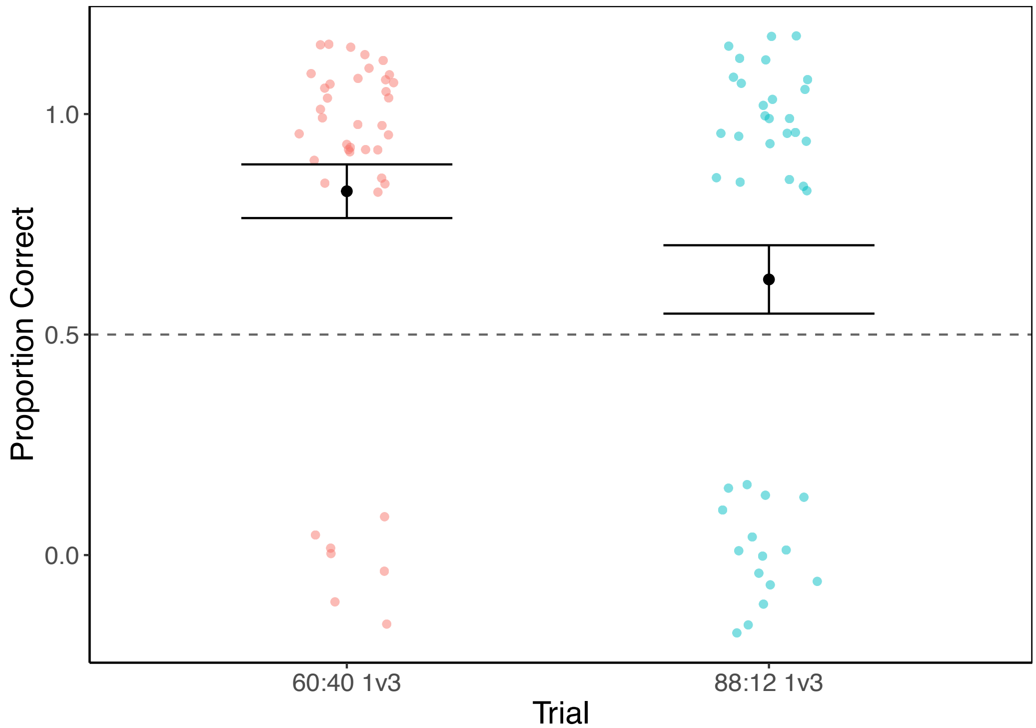

Figure 14. Mean correct responses on each trial indicated by dot. Error bars indicate ± 1 standard error of the mean and individual data points are jittered for visibility. The dashed line represents chance (50%) performance.

<u>Discussion</u>

This experiment demonstrates perhaps the first evidence of toddlers integrating probability and number. It appears that beginning in the second year of life, but likely not before then, children can consider both the distributions and the number of draws when reasoning about sampling and probability.

While the results were not significantly above chance on both types of trials, unlike infants, toddlers showed flexibility in choosing the cup with more draws more often when it was more advantageous (the 60:40 1v3 trial) than when it was less advantageous (the 88:12 1v3 trial). To look at this in another way, I coded the responses for choice whereby picking the 1-lollipop draw was coded as a 0 and picking the 3-lollipop draw was coded as a 1. The tendency to choose the three-item draw more often on the 60:40 1v3 trial than the 88:12 1v3 trial was statistically significant (*Wald* $X^2(df = 1, N = 40) = 27.39, p < .001$). This main effect found for choosing the larger draw across trials indicated that they integrated numerical and proportional information such that one choice is preferred in each trial type. In other words, there appears to be some sensitivity to the shifting value of choosing each cup based on the strength of the numerical and proportional information provided.

Interestingly, toddlers performed better on the more challenging trial, statistically speaking. This is a puzzling finding, but there are two potential reasons for this. I speculate that perhaps not having immediate access to the proportional information at the time of making a choice may have led some of the toddlers to be drawn to the larger number of lollipops in the sample. It is possible that in the 60:40 1v3 trial, where the proportions appear close to 50:50, toddlers' did not detect the difference between the ratios and reasoned that, all else equal, more draws is better. However, in the 88:12 1v3 trial, they start out with clearly skewed distributions where they may initially be drawn to the container with a greater proportion of target objects but then seeing the additional draws from the distribution with the lower proportion of target objects introduces uncertainty and they cannot quite compute how or whether this should change their decision. Additionally, it may be the case that while having visible distributions with older children in Chapter 2 (Experiment 1 & 2) caused an over-weighting of proportional information due to its salience, *hiding* the proportional information may have been detrimental for toddlers. The working memory capacities and limitations for proportional information specifically is not known for this age range, and therefore it may be the case that the distributions should have remained visible in this experiment. An interesting and necessary follow up for this work would be to examine working memory demands for the ratio information in this age range specifically.

This would involve a closer investigation of whether or not toddlers detect the difference in these proportions and if they remember the proportions well enough to integrate them with additional information.

As mentioned in the introduction to this chapter, there is little research investigating the cognitive development of probabilistic decision making based on numerical information in the age range between infancy and preschool years. In this chapter I examined these abilities and found that under some condition, toddlers may be capable of combining information about the proportion of a distribution and the number of sampled objects to make probabilistic decisions. This may resemble the beginnings of developing some sensitivity to the value placed on different kinds of numerical information. Next, I move on to a broader discussion of the findings across my entire dissertation, the limitations, future directions and broad implications of this work in Chapter 5.

Chapter Five: General Discussion

Summary of Findings

This dissertation outlined my investigation of the use of numbers and proportions in probabilistic decision-making across different stages of development. Chapter 2 examined these abilities in school-aged children (5- and 6-year-olds) across four experiments, and a fifth experiment extended this investigation to adults. The findings from the first experiment revealed that children are able to track probabilities fairly well when making judgements about sampled objects from larger distributions. Of particular interest was the children's chance performance when there were identical distributions containing an equal number of target to non-target objects. This first experiment also brought our attention to the fact that task demands and the idiosyncrasies of the instructions used may potentially play a role in children's performance by altering their understanding of the goal of the task, and subsequently impacting the strategy they use to respond to each trial. In Experiment 2, which featured multi-object samples and identical proportional properties of the distributions to Experiment 1, children showed improved performance, but were still struggling with the most challenging trial, 60:40 2v5. In Experiment 3, I examined the possibility that the proportional information of the distributions was more salient than the numerical information of the samples, and modified the design slightly to occlude the distributions at choice. Results showed that accuracy was significantly higher than chance for all four trial types, suggesting that there was some merit to the salience of the distributions idea. Experiment 4 was both a partial replication of the 88:12 1v3 and 60:40 1v3 trials in Experiment 1, except with occluded distributions at choice, as well as an investigation of baseline proportional discrimination abilities by adding an 88:12 1v1 and a 60:40 1v1 trial. Children succeeded on the 88:12 1v3 trial and still struggled with the 60:40 1v3 trial, but they succeeded on both trial types when samples were equal, choosing the single draw that came from the distribution with a higher proportion of target objects, indicating that they were able to accurately reason about the proportional information presented to them. Finally, in Experiment 5, adults were tested on all trial types presented to children in Experiments 1-4 and succeeded on all of them with the exception of the 60:40 2v5 trial where their accuracy was no different from chance.

Chapters 3 and 4 extended this work to infants and toddler, respectively using a physical version of the task. I found that infants were, as a group, responding at chance levels for both trial types used. Toddlers showed flexibility in choosing the cup with more draws more often when it was advantageous (the 60:40 1v3 trial) than when it was less advantageous (the 88:12 1v3 trial). While the results were not significantly above chance on both types of trials, unlike with infants, a main effect found for choosing the larger draw across trials indicated that there is some sensitivity to the shifting strength of the numerical and proportional information provided. This provided promising evidence with regard to performance on tasks that require the integration of numerical and proportional information, but also leads to many open questions regarding the mechanisms that may be responsible for these patterns of results.

Conceptual Implications

Cross talk between large and small number systems?

The tasks used in Chapters 2-4 of this dissertation require that participants combine proportional information about the distributions, likely represented by the ANS, and small numbers of the sampled objects. This raises the question of whether or not success in this task requires some form of cross-talk between these two distinct systems. There are three theoretical perspectives on whether or not there is cross-talk between the large and small number systems.

The first of these perspectives is that the two systems are independent of one another and that there is no cross-talk between them (Feigenson et al., 2002; vanMarle, 2013). The second is that the ANS might be used for both small numbers as well as large numbers, thus requiring no cross-talk among systems (Cordes et al., 2001; Starr et al., 2013). And finally, the two systems may be independent of one another, but cross-talk between the two may be possible under certain circumstances that would allow for reasoning about both small and large numbers simultaneously.

There are some open questions about why infants failed on this task, but also about how toddlers might have succeeded. One possibility for this pattern of results is a development of the capacity to combine information from both sources. While there may be evidence that adults can use the ANS to represent both small and large numbers (Cordes et al., 2001), the work by Feigenson and colleagues suggests that this is not necessarily the case for infants. If infant reasoning about small numerosities employed the ANS then one would expect to see the same ratio-dependent discrimination despite the actual values compared. This, however, is not the case, as they perform at chance when presented with a 1 vs. 4 comparison, or a 2 vs. 4 comparison (Feigenson, & Carey, 2003, 2005). This lends support to the notion that infants use two different systems to represent these numbers and reason about them, one that has an upper limit used for small numerosities under 4, and another that is used for approximating large numerositites in a ratio-dependent fashion. This upper limit on infants' ability to track small numbers of objects lends support to the notion that parallel individuation is recruited and that it is distinct from the system responsible for representing large numbers (e.g. Feigenson et al., 2002). Preschoolers between the ages of two-and-a-half and four-and-a-half are also reliably more accurate in comparing sets of small numerosities (2 vs. 3) than pairs of numerosities larger than 5 with the same ratio, even when controlling for continuous variables such as surface area and total perimeter and presented too briefly for counting (Cheung & LeCorre, 2018). While this difference in performance can be interpreted to indicate that preschool-aged children use distinct systems to compare small and large numerosities, this remains an open question because it is unknown if the ANS can be used for both (perhaps with reduced acuity) at this age.

Other studies, however, have suggested that in some contexts, infants may be able to use the ANS to represent small numerosities in addition to large ones (Cordes & Brannon, 2009b; Starr et al., 2013). In a numerical change detection paradigm that does not involve habituation or familiarization to a single numerical value but instead requires infants to detect numerical change in one visual stream, 6-month-old infants were tested with different combinations of small number sets (e.g. 1v2, 1v3, 2v3) to determine whether the same ratio-dependent discrimination signature that is a hallmark of the ANS and discriminating between large numerosities exists for small ones. Infants successfully discriminated 1 versus 3 and 1 versus 2, but failed with the 2 versus 3 comparison (Starr et al., 2013). If the small number system was being used, infants should have succeeded at discriminating 2 versus 3 objects. Following this up with a test of comparing numbers that cross the small to large divide, infants succeeded in discriminating between 2 and 4 objects, suggesting that it is likely one system that is at play, and one with a ratio dependence that is very similar to that of the ANS. The same ratio discrimination signature was found for small number sets as well as large number sets that both differed by the same ratio, mirroring results found for infants discriminating between large numerosities that differ by a ratio of 1:2 but not 2:3 (Xu & Spelke, 2000). This suggests that the ANS can be recruited to detect numerical change despite absolute numerical values. These results can be interpreted as evidence in support of the notion that the ANS can be recruited to represent numerosities regardless of set size, at least using the numerical change detection paradigm and some auditory habituation studies, and that factors such as whether participants have access to their full attentional resources, and whether the spacing between individual objects is within the spatial resolution of attention may affect whether the ANS or parallel individuation is employed (Starr et al., 2013; vanMarle & Wynn, 2009).

Proponents of the view that there is no cross-talk between the system for representing small numbers and the system for representing large numbers are supported by the findings that infants fail to discriminate sets of small numbers from sets of large numbers (Feigenson et al., 2002; vanMarle, 2013). Other studies have shown that this can be done under some circumstances, but it requires a higher ratio between the numbers being compared to do so effectively. When comparing small number sets to other small number sets, infants can discriminate between two numbers that differ by a ratio of 1:2, whereas if the two sets being compared contain one large number (>3) and one small one (<4), this sensitivity threshold is increased to 1:4 (Cordes & Brannon, 2009b). This is interpreted by some to indicate that one system is responsible for this comparison that crosses the small-large number boundary, but another plausible hypothesis is that the two systems *can* in fact be integrated to simultaneously represent small and large numbers and be able to compare them, but that this is a difficult strategy that results in markedly reduced accuracy. This reduced accuracy that is the result of the noisy cross-talk between the two systems may even become a little more refined across development.

The findings in this dissertation offer insights into the mechanisms used for integrating two sources of numerical information. However, there still remains much to uncover with respect to the strategies that would allow for success on this task. If true integration of proportional information and sample size occurs in order to reliably select the option that would yield the higher probability of getting target objects, one of two potential mechanisms must be employed. The first would be that information about the sample size is processed separately from information about the population distributions. The proportion of target to non-target objects is then computed for that distribution, then the information from these two representational systems are somehow assimilated, and finally, the two computed probabilities of each population-sample set are compared. This complex process implies that there is some level of cross-talk between the two systems for representing small and large numbers. The alternative account would be that one universal system is capable of supporting these representations and computations – the ANS. Because the ANS is involved in representing different quantities across modalities in a commensurate manner, it may facilitate the integration of these different sources of numerical information. It is important to consider that there may be changes occurring in use of these systems with age, which may also lend support to the notion that the ANS is involved in these integrative processes, but become more adept at doing so later in toddlerhood.

Issue of Continuous Variables

One other key issue that warrants discussion is how to tell if participants are reasoning about discrete number rather than any other feature of the visual stimuli we present them. For instance, confounding volume or overall area with numerosity introduces unanswered questions regarding the mechanism of representing these objects and engaging in comparative reasoning about them. While there is evidence that infants may actually be tracking multiple different features of a set, including numerosity as well as other continuous variables like contour or overall area (Cordes & Brannon, 2009a), many studies of small number reasoning and discrimination confound numerosity with other continuous variables (e.g., Clearfield & Mix, 2001; Feigenson et al., 2002). When infants are habituated to either 1 or 2 objects (Feigenson et al., 2002) or 2 or 3 objects (Clearfield & Mix, 2001) of a constant size and then tested with either the familiar number of objects with a novel size or a novel number of objects with the familiar size they dishabituate only to changes in object size and not to changes in object number. Even more interesting is that in some studies, when continuous variables such as surface area are controlled, infants no longer show discrimination of novel numerosities (Feigenson et al., 2002). Similarly, when continuous extent is varied in a habituation paradigm, 6-month-olds fail to discriminate 1 versus 2 and 2 versus 4 dots (Xu, 2003; Xu et al., 2005). This is in direct contrast with evidence found in other work that suggests that perhaps the ANS is involved in small number comparisons as well (Starr et al., 2013). One potential explanation, however, is that different paradigms and testing conditions have different working memory demands that impact performance. It may not be the case that infants' can no longer represent 4 items and compare them to another set using their small number system, but rather that this reaches the threshold of working memory for them and they can no longer track that many individual objects. Loading the infants' memory may prevent them from tracking objects using their small number system, bypassing the upper limit effect. It is an open question as to whether participants could have been using features that correlate with number rather than numbers per se in this dissertation, but future work could try to tease this apart by manipulating things like item size in the displays, for example.

The experiments in this dissertation were not designed with careful controls of variables that are correlated with both the samples and the proportions. This leaves open the question of whether participants of any age abstracted the ratios based on number and represented the sampled items based on number, and not other, correlated variables. When two displays are presented in a comparative judgment task, they involve not only numerosity but also other visual properties: average diameter, total surface (the sum of the surface area of all the dots in each array), convex hull (the smallest contour that includes all of the dots), density (area extended divided by total surface), and total circumference (the sum of the circumferences of all the dots in an array). Thus, many have argued that when asking participants to judge which of two arrays has more objects (e.g. dots), it is possible that they rely on, or are to a certain extent affected by, the continuous non-numerical properties in the display. Attempting to control for these continuous variables that may confound investigations of number reasoning alone is important, but has also proven nearly impossible (for a review, see Henik et al., 2017). It is difficult to control for continuous properties that confound numerosity because the various visual properties correlate imperfectly. For example, when one changes the convex hull, one changes the density of the dots, but the change in density cannot be determined perfectly by the change in the convex hull. This is true for the contingencies among all visual properties, and thus it is very complicated (if not impossible) to control all properties at once (Gevers et al., 2016; Leibovich et al., 2017; Mix et al., 2002; Núñez, 2017).

Information Integration

There are a number of broad questions that remain open for discussion and require further consideration and future exploration. Firstly, the nature of the tasks I used do not allow a complete and in-depth understanding of the specific strategy-use that is at play in contexts where individuals need to consider both proportional and sample numbers to make a decision under uncertainty. It is unclear if the predominant strategy that is used is one where one piece of information is evaluated, for example, then the other piece of information is used to update that judgement, either confirming or changing the initial conclusion. And if so, do individuals tend to process proportional information of the distributions first, then assess the sample number, or could they use the proportional information to update their decision based on number? More crucially, how do these processes differ when the two pieces of information are conflicting, or point in opposite directions, versus when they both point to the same outcome. For example, one can imagine that it can be challenging in cases where the proportional information initially makes one option appear favorable, only to have to update this evaluation when the sample size information alters this favorability, which is the case for all of the test trials used in this dissertation. Another potential mechanism for decision-making under these circumstances is a combinatorial one. This account would involve a combining of each set of proportion/sample-size pairs separately first before comparing them to one another.

While this sounds very complex and challenging, we know from other work on information integration (reviewed in Chapter 2) that children from a relatively young age are capable of engaging in this kind of information combining in decision-making. In addition to the work reviewed in Chapter 2 on the integration of numerical information, research on children's abilities to integrate multiple pieces of information in decision-making extends to the social domain as well (Bridgers et al., 2016). In a blicket machine task where children were told that some shapes are blickets that activate a machine, 4-year-olds were provided with both testimony information as well as causal frequency information. This involved being presented with testimony from either a confident informant or an uncertain one, and then being presented with a sequence of events where a shape activates the blicket machine either deterministically (100% of the time) or probabilistically (e.g. 66% of the time). When the testimony information conflicts with the sequence of events, children show evidence of considering both sources of information when making judgements about which shape is the blicket. This therefore supports the notion that children as young as 4 are able to integrate observed causal frequency information with testimony information. Similarly, children as young as 4 years of age show sensitivity to both witness testimony as well as proportional information when presented with both sources of information, whereby they rely more on proportional information when the witness is unreliable, but disregard proportional information if the conflicting testimony comes from a reliable witness (Gualtieri et al., 2020).

There is also evidence that starting at the age of 4, children use prior knowledge to inform and limit the hypothesis space surrounding inferences made under uncertainty. In particular, children take into consideration the base-rate proportions of target options in their judgement of whether or not they require more information to make a decision. Three- and 4-year-olds were tested in a blicket-machine paradigm and were first given a training phase in which the frequency of blickets was varied such that blickets were either rare or common (Sobel et al., 2004). They were then tested on whether or not they used this information to guide their inferences about uncertain information. When an object displayed unambiguous causal function, all children judged it to be a blicket, regardless of whether blickets were rare or common. However, when another object was ambiguous in whether or not it was causally functional, 4-year-olds were more likely to determine that they were blickets if blickets were common than if they were rare. 3-year-olds on the other hand did not show sensitivity to this manipulation of prior probabilities.

Limitations

This dissertation has a number of limitations. First, in continuation of the discussion on the importance of controlling for non-numerical variables in all work that aims to investigate numerical reasoning in any capacity, I acknowledge the lack of careful controls of continuous variables in these studies. these experiments are a first step to see if kids can do a very difficult integration task when all cues (numerical and correlated non-numerical cues) point in the same direction. Future work can examine whether these findings would hold up with more careful controls. In order to do this, I conflated total visible surface area with number when I presented the population distributions of lollipops. Also, regarding lack of perfectly controlled stimuli, because I used physical lollipops that could move around in Chapters 3 and 4, each individual participants' view of the distribution might have been slightly different depending on what was visible after shaking the containers around

Other limitations of these experiment include things that can and should be done differently in future work in order to provide a more rigorous study of numerical integration in probabilistic decision-making. The first would be that none of the experiments in this dissertation have manipulated the visibility of proportions within-subjects. I started off with Experiments 1 and 2 in Chapter 2 that presented participants with visible proportions throughout the course of the task, including the time at which participants made their choices. The rest of the experiments that follow have an occlude that hides the proportions at the time of sampling from the distributions and when the participant is choosing one of the two outcomes. I made speculations with regard to the effect of this change in the design, but would allow for a better analysis of the effect of this manipulation and an ability to draw actual conclusions about its effect had this been manipulated in a within-subjects design.

All of the experiments in this dissertation also failed to reveal any age effects or age by trial interactions. While it may very well be the case that no development occurs in these abilities at these various developmental stages, it could also be the case that my selected age ranges were too narrow to detect age effects or too few participants were tested, especially in the infant and toddler experiments. This limitation of not testing a broader age range and larger number of participants may have masked age differences indicative of developmental changes.

Additionally, the dichotomous variable used that only assesses correct and incorrect responses by analyzing group means is a relatively coarse approach, and it in some ways limits the information gained from each participant (as opposed to other dependent variables that might produce more fine-grained data). A group mean that is no different than chance levels could indicate that children as a group were guessing, but it could instead point to different children employing different strategies, which was impossible to detect in my experiments. One way to address this in future work would be that, given that we have this programmed in Qualtrics for adults, this can also be a good candidate for online studies with children where we might be able to present a larger number of trials. We can test them with a much larger battery of problems where proportions and draw numbers are manipulated systematically. This would allow a more in-depth analysis of individual response patterns, and can help us answer the question of whether some children rely more on number, while others may rely more on proportions, and what kinds of probabilities and numbers tip the scale for these children individually. It might also allow us to ask children for scaled responses or confidence judgements when deciding between the two sample containers.

Future directions

There are a number of directions that future work can explore in this field and in the area of integrating multiple sources of numerical information in the context of probabilistic decision-making. We need to address the question of why children succeeded on the challenging 60:40 2v5 trial in Experiment 3, while adults failed on this same trial type in Experiment 5. A simple explanation would be that children simply forgot about the proportional information, or thought it was close enough to 50:50, and were influenced by the many draws in the 5-item sample taken from the distribution with the slightly lower proportion of target objects. This explanation is not straightforward, however, in light of their poor performance on the 60:40 2v5 trial in Experiments 1, 2, and 4. Another potential reason could be that there is a decrement caused by the summary sampling that the adults experienced, as opposed to the sequential, experienced sampling that the children observed. Work investigating these two types of sampling in the context of probabilistic decision making has coined the term "description-experience gap" to describe the discrepancies between decisions based on observed statistical information as opposed to summary descriptions of this information (Hertwig et al., 2004; Hau et al., 2010). This is should be further examined in future work by investigating the effect on performance of different kinds of sampling methods.

A necessary follow-up to the infant work in Chapter 3 that is particularly intriguing would be to employ eye-tracking technology to design a looking-time version of this kind of paradigm. Besides the obvious justification that this would help us overcome additional demands added on by the explicit choice paradigm, it would also allow for the presentation of many more trials which can produce continuous data instead of just two trials with a right/wrong response. It would allow for a better understanding of infants' allocation of attention in this task. Using eye-tracking, we would be able to have a quantitative measure of looking time and fixation on different areas of interest, as well as the duration of fixation for each of those. This would be informative in the amount of time spent on each of the different sources of information provided. Fixation sequences and data on the temporal order of those fixations will also shed light on what infants look at first, and what order they consider all the data in. Would they look first at the proportions of the distributions to abstract the ratio information, then at the sampled objects before making a choice? Or would there be a pattern of back-and-forth oscillation between fixating on the proportional information and the sample numbers? Do they consider each sample-proportion pair individually and compare them to one another, or do they compare proportions first before moving on to compare sample numbers? Anticipatory looking may also be used here. This would be especially helpful in assessing infants' initial expectations about which sample an agent should choose in a third person version of this task, without having to actually crawl to the cup. There may be a discrepancy between implicit measures and the explicit behaviour of crawling or even pointing to one of two choices. It may be the case that some infants and/or toddlers wish to further explore the sampled objects themselves, and that their action to move towards a particular cup does not accurately map on to their understanding that *this* would be the cup with more target objects.

Conclusions

Much is yet unknown about the development of numerical integration in probabilistic contexts. Using multiple sources of information effectively under conditions of uncertainty to maximize favorable outcomes is a complex and intricate process, and yet we do it almost every day, in many different aspects of our decision-making existence. This dissertation is an attempt at uncovering some of the patterns in these kinds of decisions and how these patterns may shift throughout development. Results show promise with regard to how well adults and school-aged children can perform on these sometimes mathematically challenging problems, indicating that there may be an intuitive nature to this kind of reasoning. Results obtained from infants highlight the need for more work to be done in this field, and with this particular age group, to better document these abilities and their strategy use. Finally, findings obtained from toddlers indicate encouraging evidence that points to the earliest age at which a sensitivity to different sources of numerical information affecting decision-making may emerge.

References

Ahl, V. A., Moore, C. F., & Dixon, J. A. (1992). Development of intuitive and numerical proportional reasoning. *Cognitive Development, 7*, 81–108. doi: 10.1016/0885-2014(92)90006-D

Antell, S., & Keating, D.P. (1983). Perception of numerical invariance in neonates. *Child Development, 54*, 695–701. doi: 10.1111/j.1467-8624.1983.tb00495.x

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

Balci, F., Freestone, D., Gallistel, C.R. (2009). Risk assessment in man and mouse. *Proceedings of the National Academy of Sciences, 106*, 2459-2463.

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 86*(3), 201–221. doi:10.1016/S0010-0277(02)00178-6

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., and Spelke, E. (2006). Nonsymbolic arithmetic in adults and young children. *Cognition, 98,* 199-222. doi:10.1016/j.cognition.2004.09.011

Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences, 102*(39), 14116–14121. doi:10.1073/pnas.0505512102

Bayless, S. & Schlottman, A. (2010). Skill related uncertainty and expected value in 5- and 7-year-olds. *Psicologica. Special Issue on Functional Measurement. 31*(3), 677-687.

Boyer, T. W., & Levine, S. C. (2015). Prompting children to reason proportionally: Processing discrete units as continuous amounts. *Developmental Psychology, 51*, 615–620. doi: 10.1037/a0039010

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology, 44*(5), 1478–1490. doi: 10.1037/a0013110

Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition, 83*, 223-240. doi:10.1016/S0010-0277(02)00005-7

Brannon, E. M., & Van de Walle, G. (2001). Ordinal numerical knowledge in young children. *Cognitive Psychology, 43*, 53–81. doi:10.1006/cogp.2001.0756

Bridgers, S., Buchsbaum, D., Seiver, E., Gopnik, A., & Griffiths, T. L. (2016). Children's causal inferences from conflicting testimony and observations. *Developmental Psychology, 54*, 9-18. doi: 9.10.1037/a0039830

Bryant, P., & Nunes, T. (2012). *Children's understanding of probability: A literature review (full report)*. London: Nuffield Foundation.

Carey, S. (2009). The origin of concepts. Oxford: Oxford University Press.

Cheung, P., & Le Corre, M. (2018). Parallel individuation supports numerical comparisons in preschoolers. *Journal of Numerical Cognition, 4,* 380-409. doi: 10.5964/jnc.v4i2.110

Choo, H., & Franconeri, S. L. (2014). Enumeration of small collections violates Weber's Law. *Psychonomic Bulletin & Review, 21*, 93-99. doi:10.3758/s13423-013-0474-4

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science, 10*, 408-411. doi:10.1111/1467-9280.00177

Cooper, R. G. (1984). Early number development: discovering number space with addition and subtraction. In C. Sophian (Eds.), *Origins of cognitive skills* (pp. 157–192). Hillsdale, NJ: Erlbaum.

Cordes, S., & Brannon, E.M. (2008). Quantitative competencies in infancy. Developmental *Science, 11* (6), 803–808. doi: 10.1111/j.1467-7687.2008.00770.x

Cordes, S., & Brannon, E.M. (2009a). The relative salience of discrete and continuous quantities in infants. *Developmental Science, 12* (3), 453–463. doi: 10.1111/j.1467-7687.2008.00781.x

Cordes, S., & Brannon, E.M. (2009b). Crossing the divide: Infants discriminate small from large numerosities. *Developmental Psychology, 45*(6), 1583-1594. doi:10.1037/a0015666

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin and Review, 8*, 698–707.

Dehaene, S. (1997). The number sense. New York, NY, USA: Oxford University Press.

Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences, 1156*, 232-259. doi: 10.1111/j.1749-6632.2009.04469.x

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 fourteen-month-old infants can predict single-event probability with large set sizes. *Developmental Science, 13*, 798–803. doi: 10.1111/j.1467-7687.2009.00943.x

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition, 130*, 335-347. doi: 10.1016/j.cognition.2013.12.001

Denison, S. & Xu, F. (2019) Infant statisticians: The origins of reasoning under uncertainty. *Perspectives on Psychological Science, 14,* 499-509. doi: 10.1177/1745691619847201

Drucker, C. B., Rossa, M. A., & Brannon, E. M. (2016). Comparison of discrete ratios by rhesus macaques (Macaca mulatta). *Animal Cognition, 19*, 75–89. doi: 10.1007/s10071-015-0914-9

Eckert, J., Call, J., Hermes, J., Herrmann, J., & Rakoczy, H. (2018). Intuitive statistical inferences in chimpanzees and humans follow Weber's Law. *Cognition, 180*, 99–107. doi: 10.1016/j.cognition.2018.07.004

Eckert, J., Call, J., & Rakoczy, H. (2017). Are great apes able to reason from multi-item samples to populations of food items. *American Journal of Primatology, 79*(10), Article e22693. doi: 10.1002/ajp.22693

Falk, R., Yudilevich-Assouline, P., & Elstein, A. (2012). Children's concept of probability as inferred from their binary choices – revisited. *Educational Studies in Mathematics, 81*(2), 207-233. doi: 10.1007/s10649-012-9402-1

Feigenson, L., & Carey, S. (2003). Tracking individuals via object files: Evidence from infants' manual search. *Developmental Science, 6*(5), 568–584. doi:10.1111/1467-7687.00313

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*(1), 295-313. doi:10.1016/j.cognition.2004.09.010

Feigenson, Carey, & Hauser. (2002). The representations underlying infants' choice of more: object files versus analog magnitudes. *Psychological Science, 13*, 150–156. doi:10.1111/1467-9280.00427

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology, 44*, 33-66. doi:10.1006/cogp.2001.0760

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314. doi:10.1016/j.tics.2004.05.002

Feigenson, L. & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition, 91,* 173-190. doi: 10.1016/j.cognition.2003.09.003

Gallistel, C. R. & Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends in Cognitive Sciences, 4*, 59–65. doi:10.1016/S1364-6613(99)01424-2

Gallistel, C. R., & Gelman, R. (2005). Mathematical cognition. In K. Holyoak & R. Morrison (Eds.). *The Cambridge handbook of thinking and reasoning* (Vol. 23, pp. 559–588). Cambridge, UK: Cambridge University Press.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review, 107,* 289-344.

Gebuis, T., Cohen Kadosh, R., & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychologica, 171*, 17–35. doi: 10.1016/j.actpsy.2016.09.003

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, Mass: Harvard University Press.

Gevers, W., Cohen Kadosh, R., & Gebius, T. (2016). Sensory integration theory: An alternative to the approximate number system. *In Continues Issues in Numerical Cognition: How Many or How Much* (Henik, A., ed.), pp. 405–418, Elsevier. doi: 10.1016/B978-0-12-801637-4.00018-4

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

Girotto, V., & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition, 106*, 325–344. doi: 10.1016/j.cognition.2007.02.005

Gualtieri, S., Buchsbaum, D., & Denison, S. (2020). Exploring information use in children's decision-making: Base-rate neglect and trust in testimony. *Journal of Experimental Psychologhy,149,* 1527-1536.

Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical experience: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making*, *23*, 48–68. doi: 10.1002/bdm.665

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*, 1457–1465. doi:10.1037/a0012682

Henik, A., Gliksman, Y., Kallai, A & Leibovich, T. (2017) Size perception and the foundation of numerical processing. *Current Directions in Psychological Science 26*(1), 48–51. doi: 10.1177/0963721416671323

Huntley-Fenner, G. & Cannon, E. (2000). Preschoolers' magnitude comparisons are mediated by a preverbal analog mechanism. *Psychological Science, 11,* 147-152. doi: 10.1111/1467-9280.00230

Hurst, M. A., & Cordes, S. (2018). Attending to relations: Proportional reasoning in 3- to 6-year-old children. *Developmental Psychology, 54*, 428–439. doi: 10.1037/dev0000440

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*(8), 534–539. doi: 10.1111/j.0956-7976.2004.00715.x

Huttenlocher, J., Duffy, S., & Levine, S. (2002). Infants and toddlers discriminate amount: Are they measuring? *Psychological Science, 13*, 244–249. doi: 10.1111/1467-9280.00445

Huttenlocher, J., Newcombe, N., & Vasilyeva, M. (1999). Spatial scaling in young children. *Psychological Science, 10*, 393–398. doi: 10.1111/1467-9280.00175

Hyde, D. C. & Spelke, E. S. (2011). Neural signatures of number processing in human infants: Evidence for two core systems underlying numerical cognition. *Developmental Science, 14,* 360-371. doi:10.1111/j.1467-7687.2010.00987.x

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. Proceedings of the National Academy of Sciences, 106(25), 10382–10385. 10.1073/pnas.0812142106

Jeong, Y., Levine, S. C., & Huttenlocher, J. (2007). The development of proportional reasoning: Effect of continuous versus discrete quantities. *Journal of Cognition and Development, 8*, 237–256. doi: 10.1080/15248370701202471

Kahneman, D., Treisman, A., & Gibbs, S. (1992). The reviewing of object-files: Object specific integration of information. *Cognitive Psychology, 24*, 175–219. doi: 10.1016/0010-0285(92)90007-O

Kahneman, D. & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11,* 123-141. doi: 10.1016/0010-0277(82)90022-1

Kayhan, E., Gredeback, G., & Lindskog, M. (2018). Infants distinguish between two events based on their relative likelihood. *Child Development, 89,* e507-e519. doi: 10.1111/cdev.12970

Kobayashi, T., Hiraki, K., Mugitani, R., and Hasegawa, T. (2004). Baby arithmetic: one object plus one tone. *Cognition, 91*: B23–B34. doi: 10.1016/j.cognition.2003.09.004

Kobayashi, T., Hiraki, K., and Hasegawa, T. (2005). Auditory-visual intermodal matching of small numerosities in 6-month-old infants. *Developmental Science 8*(5): 409–19. doi: 10.1111/j.1467-7687.2005.00429.x

Koechlin, E., Dehaene, S., and Mehler, J. (1997). Numerical transformations in five-month-old human infants. *Cognition, 3*, 89–104. doi: 10.1080/135467997387425

Lawson, C. A., & Rakison, D. H. (2013). Expectations about single event probabilities in the first year of life: The influence of perceptual and statistical information. *Infancy, 18*, 961–982. doi: 10.1111/infa.12014

Le Corre, M. & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*, 395-438. doi:10.1016/j.cognition.2006.10.005

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From 'sense of number' to 'sense of magnitude' – The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences*, 1–62. doi: 10.1017/S0140525X16000960

Libertus M. E. & Brannon, E. M. (2009). Behavioral and neural basis of number sense in infancy. *Current Directions in Psychological Science, 18,* 346-351. doi: 10.1111/j.1467-8721.2009.01665.x

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large number discrimination in human infants. *Psychological Science, 14*, 396 – 401. doi:10.1111/1467-9280.01453

Lipton, J. S., & Spelke, E. S. (2004). Discrimination of large and small numerosities by human infants. *Infancy, 5*, 271 – 290. doi: 10.1207/s15327078in0503_2

Margolis, E. (2020). The Small Number System. *Philosophy of Science, 87*(1), 113–134. doi:10.1086/706087

Matthews, P. G., & Chesney, D. L. (2015). Fractions as percepts? Exploring cross-format distance effects for fractional magnitudes. *Cognitive Psychology, 78*, 28–56. doi: 10.1016/j.cogpsych.2015.01.006

McCall, R. B. & Kagan, J. (1967). Stimulus-schema discrepancy and attention in the infant. *Journal of Experimental Child Psychology, 5*(3), 381-390. doi: 10.1016/0022-0965(67)90066-5

McCrink, K., & Spelke, E. S. (2010). Core multiplication in childhood. *Cognition, 116*, 204–216. doi: 10.1016/j.cognition.2010.05.003

McCrink, K., & Spelke, E. S. (2016). Non-symbolic division in childhood. *Journal of Experimental Child Psychology, 142*, 66–82. doi: 10.1016/j.jecp.2015.09.015

McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science, 15*, 776–781. doi:10.1111/j.0956-7976.2004.00755.x

McCrink, K., & Wynn, K. (2007). Ratio Abstraction by 6-Month-Old Infants. *Psychological Science, 18*(8), 740–745. doi:10.1111/j.1467-9280.2007.01969.x

Mix, K.S., Huttenlocher, J., & Levine, S.C. (2002). Quantitative development in infancy and early childhood. Oxford University Press; New York, NY.

Moher, M., Tuerk, A.S., & Feigenson, L. (2012). Seven-month-old infants chunk items in memory. *Journal of Experimental Child Psychology, 112,* 361-377. doi:10.1016/j.jecp.2012.03.007

Nuñez, R. E. (2017). Is There Really an Evolved Capacity for Number? *Trends in Cognitive Sciences, 21*(6), 409–424. doi: 10.1016/j.tics.2017.03.005

O'Grady, S., Griffiths, T. L., & Xu, F. (2016). Do simple probability judgments rely on integer approximation? In C. Kalish, M. Rau, J. Zhou, & T. Rogers (Eds.), *Proceedings of the 40th annual conference of the cognitive science society* (pp.1008–1013). Philadelphia, PA: Cognitive Science Society.

O'Grady, S. & Xu, F. (2020). The development of nonsymbolic probability judgements in children. *Child Development, 91,* 784-798. doi: 10.1111/cdev.13222

Odic, D. (2017). Children's intuitive sense of number develops independently of their perception of area, density, length, and time. *Developmental Science, 21*(6). doi: 10.1111/desc.12533

Odic, D., Libertus, M., Feigenson, L., & Halberda, J. (2013). Developmental Change in the Acuity of Approximate Number and Area Representations. *Developmental Psychology, 49,* 1103-1112. doi:10.1037/a0029472

Piaget, J. (1965). *The child's conception of number*. Norton; New York, NY.

Piaget, J., & Inhelder, B. (1975). *The origins of the idea of chance in children*. New York, NY: Norton & Company.

Pylyshyn, Z.W. (1989). The role of location indexes in visual perception: A sketch of the FINST spatial index model. *Cognition, 32*, 65–97. doi: 10.1016/0010-0277(89)90014-0

Pylyshyn, Z.W. (1994). Some primitive mechanisms of spatial attention. *Cognition, 50*, 363–384. doi: 10.1016/0010-0277(94)90036-1

Rakoczy, H., Clüver, A., Saucke, L., Stoffregen, N., Gräbener, A., Migura, J. & Call, J. (2014). Apes are intuitive statisticians. *Cognition, 131*(1), 60-68. doi: 10.1016/j.cognition.2013.12.011

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science, 19*, 607-614. doi:10.1111/j.1467-9280.2008.02130.x

Rosenberg, R. D. & Feigenson, L. (2013). Infants hierarchically organize memory representations. *Developmental Science, 16,* 610-621. doi:10.1111/desc.12055

Rousselle, L., & Noël, M. P. (2008). The development of automatic numerosity processing in preschoolers: Evidence for numerosity-perceptual interference. *Developmental Psychology, 44*, 544-560. doi:10.1037/0012-1649.44.2.544

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

Schlottman, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development, 72,* 103-122. doi: 10.1111/1467-8624.00268

Schlottman, A. & Anderson, N.H. (1994). Children's judgments of expected value. *Developmental Psychology, 30,* 56-66. doi: 10.1037/0012-1649.30.1.56

Schlottmann, A., & Tring, J. (2005). How children reason about gains and losses: Framing effects in judgement and choice. *Swiss Journal of Psychology, 64*(3), 153-171. doi: 10.1024/1421-0185.64.3.153

Schulze, C., & Hertwig, R. (2021). A description-experience gap in statistical intuitions: Of smart babies, risk-savvy chimps, intuitive statisticians, and stupid grown-ups. *Cognition, 210*, Article 104580. doi: 10.1016/j.cognition.2020.104580

Simon, T. (1997). Reconceptualizing the origins of number knowledge: A non-numerical account. *Cognitive Development, 12*, 349–372. doi: 10.1016/S0885-2014(97)90008-3

Singer-Freeman, K. E., & Goswami, U. (2001). Does a half pizza equal half a box of chocolate? Proportional matching in an analogy task. *Cognitive Development, 16*, 811–829. doi: 10.1016/S0885-2014(01)00066-1

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*, 303-333. doi: 10.1207/s15516709cog2803_1

Sophian, C. (2000). Perceptions of proportionality in young children: Matching spatial ratios. *Cognition, 75*, 145–170. doi: 10.1016/S0010-0277(00)00062-7

Sophian, C., & Adams, N. (1987). Infants' understanding of numerical transformations. *British Journal of Developmental Psychology, 5*, 257–264. doi: 10.1111/j.2044-835X.1987.tb01061.x

Starkey, P., & Cooper, R.G., Jr. (1980). Perception of numbers by human infants. *Science, 210*, 1033–1035. doi: 10.1126/science.7434014

Starkey, P., Spelke, E.S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36*, 97–128. doi: 10.1016/0010-0277(90)90001-Z

Starr, A., Libertus, M.E., & Brannon, E.M. (2013). Infants show ratio-dependent number discrimination regardless of set size. *Infancy, 18,* 927-942. doi:10.1111/infa.12008

Strauss, M.S., & Curtis, L.E. (1984). Development of numerical concepts in infancy. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 131–155). Hillsdale, NJ: Erlbaum.

Tecwyn, E., Denison, S., Messer, E.J., & Buchsbaum, D. (2017). Intuitive probabilistic inference in capuchin monkeys. *Animal Cognition, 20*, 243-256. doi: 10.1007/s10071-016-1043-9

Teglas, 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 of the United States of America, 104*, 19156–19159. doi: 10.1073/pnas.0700271104

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. doi: 10.1126/science.185.4157.1124

Uller, C., Huntley-Fenner, G., Carey, S., & Klatt, L. (1999). What representations might underlie infant numerical knowledge. *Cognitive Development, 14*, 1–36. doi:10.1016/S0885-2014(99)80016-1

van Oeffelen, M.P. & Vos, P.G. (1982) A probabilistic model for the discrimination of visual number. *Percept. Psychophys. 32*, 163–170. doi: 10.3758/BF03204275

vanMarle, K.(2013). Infants use different mechanisms to make small and large number ordinal judgments. *Journal of Experimental Child Psychology, 114*(1), 102-110. doi: 10.1016/j.jecp.2012.04.007

vanMarle, K., Mou, Y., & Seok, J. H. (2016). Analog magnitudes support large number ordinal judgments in infancy. *Perception, 45*(1-2), 32-43. doi: 10.1177/0301006615602630

vanMarle, K., & Wynn, K. (2009). Infants' auditory enumeration: Evidence for analog magnitudes in the small number range. *Cognition, 111,* 302-316.

vanMarle, K., & Wynn, K. (2011). Tracking and quantifying objects and non-cohesive substances. *Developmental Science, 14*(3), 502-515. doi: 10.1111/j.1467-7687.2010.00998.x

Waismeyer, A., Meltzoff, A. N., & Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: An action measure. *Developmental Science, 18,* 175-182. doi: 10.1111/desc.12208

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

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*, 130–137. doi: 10.1111/1467-9280.00120

Wood, J. N., & Spelke, E. S. (2005). Infants' enumeration of actions: Numerical discrimination and its signature limits. *Developmental Science, 8*(2), 173–181. doi:10.1111/j.1467-7687.2005.00404.x

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*, 749–750. doi: 10.1038/358749a0

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition, 89*, B15–B25. doi:10.1016/S0010-0277(03)00050-7

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 of the United States of America, 105*, 5012–5015. doi: 10.1073/pnas.0704450105

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11. doi:10.1016/S0010-0277(99)00066-9

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science, 8*, 88–101. doi:10.1111/j.1467-7687.2005.00395.x

Yost, P. A., Siegel, A. E., & Andrews, J. M. (1962). Nonverbal probability judgments by young children. *Child Development, 33*(4), 769-780. doi: 10.2307/1126888

Zosh, J. Feigenson, L. Halberda, J. (2011). Memory for multiple visual ensembles in infancy. *Journal of Experimental Psychology*, *140*, 141-158. doi:10.1037/a0022925

Appendix

Table 1. Computations for obtaining more targets on all possible comparisons for the 50:50 1v3 trial. Draw probabilities computed with replacement. G = Green (i.e., target); B= Black

50:50 1 draw	50:50 3 draws	Probability of combination
G (.5)	3 G (.125)	.0625
G (.5)	2 G ,1 B (.375)	.1875
G (.5)	1 G, 2 B (.375)	.1875
G (.5)	3 B (.125)	.0625
B (.5)	3 G (.125)	.0625
B (.5)	2 G ,1 B (.375)	.1875
B (.5)	1 G, 2 B (.375)	.1875
B (.5)	3 B (.125)	.0625
Ignoring ties, one is 11 times more likely to win choosing the 3 draws		

Table 2. Computations for obtaining more targets on all possible comparisons for the 50:50 2v5 trial. Draw probabilities computed with replacement. G = Green (i.e., target); B= Black

50:50 2 draw	50:50 5 draws	Probability of combination
2 B (.25)	5 G (.0312)	.00781
2 B (.25)	4 G, 1 B (.1563)	.0391
2 B (.25)	3 G, 2 B (.3125)	.0781
2 B (.25)	2 G, 3 B (.3125)	.0781
2 B (.25)	4 B, 1 G (.1563)	.0391
2 B (.25)	5 B (.0312)	.00781
2 G (.25)	5 G (.0312)	.00781
2 G (.25)	4 G, 1 B (.1563)	.0391
2 G (.25)	3 G, 2 B (.3125)	.0781
2 G (.25)	2 G, 3 B (.3125)	.0781
2 G (.25)	4 B, 1 G (.1563)	.0391
2 G (.25)	5 B (.0312)	.00781
1 G, 1 B (.5)	5 G (.0312)	.0158
1 G, 1 B (.5)	4 G, 1 B (.1563)	.0782
1 G, 1 B (.5)	3 G, 2 B (.3125)	.156
1 G, 1 B (.5)	2 G, 3 B (.3125)	.156
1 G, 1 B (.5)	4 B, 1 G (.1563)	.0782
1 G, 1 B (.5)	5 B (.0312)	.0158
Ignoring ties, one is 12.3 times more likely to win choosing the 3 draws		

Table 3. Computations for obtaining more targets on all possible comparisons for the 88:12 1v3 trial. Draw probabilities computed with replacement. G = Green (i.e., target); B= Black

88:12 1 draw	12:88 3 draws	Probability of combination
G (.88)	3 G (.00179)	.00158
G (.88)	2 G ,1 B (.0381)	.0335
G (.88)	1 G, 2 B (.279)	.246
G (.88)	3 B (.681)	.599
B (.12)	3 G (.00179)	.000215
B (.12)	2 G ,1 B (.0381)	.00457
B (.12)	1 G, 2 B (.279)	.0335
B (.12)	3 B (.681)	.082
Ignoring ties, one is 8.2 times more likely to win choosing the 1 draw		

Table 4. Computations for obtaining more targets on all possible comparisons for the 88:12 2v5 trial. Draw probabilities computed with replacement. G = Green (i.e., target); B= Black

88:12 2 draw	12:88 5 draws	Probability of combination
2 B (.0144)	5 G (.0000249)	3.586E-07
2 B (.0144)	4 G, 1 B (.000912)	1.313E-05
2 B (.0144)	3 G, 2 B (.0134)	0.000193
2 B (.0144)	2 G, 3 B (.0981)	0.00141
2 B (.0144)	4 B, 1 G (.360)	0.00518
2 B (.0144)	5 B (.528)	0.00760
2 G (.774)	5 G (.0000249)	1.927E-05
2 G (.774)	4 G, 1 B (.000912)	0.000706
2 G (.774)	3 G, 2 B (.0134)	0.0104
2 G (.774)	2 G, 3 B (.0981)	0.0759
2 G (.774)	4 B, 1 G (.360)	0.279
2 G (.774)	5 B (.528)	0.409
1 G, 1 B (.212)	5 G (.0000249)	5.279E-06
1 G, 1 B (.212)	4 G, 1 B (.000912)	0.000193
1 G, 1 B (.212)	3 G, 2 B (.0134)	0.00284
1 G, 1 B (.212)	2 G, 3 B (.0981)	0.0208
1 G, 1 B (.212)	4 B, 1 G (.360)	0.0763
1 G, 1 B (.212)	5 B (.528)	0.112
Ignoring ties, one is 19 times more likely to win choosing 1 draw		

Table 5. Computations for obtaining more targets on all possible comparisons for the 60:40 1v3 trial. Draw probabilities computed with replacement. G = Green (i.e., target); B= Black

60:40 1 draw	40:60 3 draws	Probability of combination
G (.6)	3 G (.064)	0.0384
G (.6)	2 G ,1 B (.288)	0.1728
G (.6)	1 G, 2 B (.434)	0.2604
G (.6)	3 B (.216)	0.1296
B (.4)	3 G (.064)	0.0256
B (.4)	2 G ,1 B (.288)	0.1152
B (.4)	1 G, 2 B (.434)	0.1736
B (.4)	3 B (.216)	0.0864
Ignoring ties, one is 4 times more likely to win choosing 3 draws		

Table 6. Computations for obtaining more targets on all possible comparisons for the 60:40 2v5 trial. Draw probabilities computed with replacement. G = Green (i.e., target); B= Black

60:40 2 draw	40:60 5 draws	Probability of combination
2 B (.16)	5 G (.0102)	0.00163
2 B (.16)	4 G, 1 B (.0768)	0.0123
2 B (.16)	3 G, 2 B (.230)	0.0368
2 B (.16)	2 G, 3 B (.346)	0.0554
2 B (.16)	4 B, 1 G (.259)	0.0414
2 B (.16)	5 B (.0778)	0.0124
2 G (.36)	5 G (.0102)	0.00367
2 G (.36)	4 G, 1 B (.0768)	0.0276
2 G (.36)	3 G, 2 B (.230)	0.0828
2 G (.36)	2 G, 3 B (.346)	0.125
2 G (.36)	4 B, 1 G (.259)	0.0932
2 G (.36)	5 B (.0778)	0.0280
1 G, 1 B (.48)	5 G (.0102)	0.00490
1 G, 1 B (.48)	4 G, 1 B (.0768)	0.0369
1 G, 1 B (.48)	3 G, 2 B (.230)	0.110
1 G, 1 B (.48)	2 G, 3 B (.346)	0.166
1 G, 1 B (.48)	4 B, 1 G (.259)	0.124
1 G, 1 B (.48)	5 B (.0778)	0.0373
Ignoring ties, one is 3 times more likely to win choosing 5 draws		