British Journal of Developmental Psychology (2018)

The British Psychological Society

www.wileyonlinelibrary.com

Special Issue Paper

Icon arrays help younger children's proportional reasoning

Azzurra Ruggeri[1,2]*, Laurianne Vagharchakian[3] and Fei Xu[4]

[1]MPRG iSearch|Information Search, Ecological and Active Learning Research With Children, Max Planck Institute for Human Development, Berlin, Germany
[2]School of Education, Technical University Munich, Germany
[3]Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany
[4]Department of Psychology, University of California, Berkeley, California, USA

We investigated the effects of two context variables, presentation format (icon arrays or numerical frequencies) and time limitation (limited or unlimited time), on the proportional reasoning abilities of children aged 7 and 10 years, as well as adults. Participants had to select, between two sets of tokens, the one that offered the highest likelihood of drawing a gold token, that is, the set of elements with the greater proportion of gold tokens. Results show that participants performed better in the unlimited time condition. Moreover, besides a general developmental improvement in accuracy, our results show that younger children performed better when proportions were presented as icon arrays, whereas older children and adults were similarly accurate in the two presentation format conditions.

Statement of contribution

What is already known on this subject?

- There is a developmental improvement in proportional reasoning accuracy.
- Icon arrays facilitate reasoning in adults with low numeracy.

What does this study add?

- Participants were more accurate when they were given more time to make the proportional judgement.
- Younger children's proportional reasoning was more accurate when they were presented with icon arrays.
- Proportional reasoning abilities correlate with working memory, approximate number system, and subitizing skills.

Many tasks in everyday life require people to understand probabilities and likelihoods to make judgements and decisions under uncertainty. Which medicine is more likely to make my stomach ache go away? Or, probably more relevant for children, how likely am I

*Correspondence should be addressed to Azzurra Ruggeri, MPRG iSearch|Information Search, Ecological and Active Learning Research With Children, Max Planck Institute for Human Development, Lenzteallee, 94, Berlin 14195, Germany (email: ruggeri@mpib-berlin.mpg.de).

DOI:10.1111/bjdp.12233

to get one of those bright red gumballs from this gumball machine? Probabilities are usually expressed as a ratio of the number of likely outcomes (e.g., red gumballs) over the total number of possible outcomes (e.g., gumballs in the machine). Sometimes people's decisions and choices involve the comparison of two ratios, which is indicated by the term *proportional reasoning*. For example, when deciding which of two gumball machines to put a quarter into, children might compare the probabilities of the two machines producing the desired red gumball. This comparison process is often approximated without calculations, as people have only limited time, computational power, and knowledge of the alternative distributions.

In this paper, we investigated the effects of two context variables, presentation format (visual or numerical), and time limitation, on the proportional reasoning skills of 7- and 10-year-old children and adults. Moreover, we explored whether and how participants' proportional reasoning skills are correlated with some of their more basic numerical and cognitive skills.

The development of probabilistic and proportional reasoning skills

Numeracy skills are correlated with education, developing, and possibly declining with age (e.g., Davids, Schapira, McAuliffe, & Nattinger, 2004; Hibbard, Peters, Slovic, & Finucane, 2001; Okan, Garcia-Retamero, Galesic, & Cokely, 2012). A growing body of research suggests that infants are already capable of probabilistic reasoning (Denison, Reed, & Xu, 2013; Denison & Xu, 2010a, 2010b; Teglas *et al.*, 2011; Xu & Denison, 2009; Xu & Garcia, 2008) and that infants and preschoolers are already able to *use* probabilistic information to form judgements, to make decisions, predictions, and generalizations, and to guide their information search (Betsch & Lang, 2013; Betsch, Lang, Lehmann, & Axmann, 2014; Denison & Xu, 2014; Gweon, Tenenbaum, & Schulz, 2010; Kushnir & Gopnik, 2005; Kushnir, Xu, & Wellman, 2010). Moreover, children are able to integrate prior probabilities with feedback and subsequent evidence (Denison, Bonawitz, Gopnik, & Griffiths, 2013; Girotto & Gonzalez, 2008; Gonzalez & Girotto, 2011) and make inferences that are consistent with the general principles of Bayesian inference (e.g., Eaves & Shafto, 2012; Gopnik & Wellman, 2012; Ruggeri, Lombrozo, Griffiths, & Xu, 2016; Schulz, Bonawitz, & Griffiths, 2007; Xu & Tenenbaum, 2007a, 2007b).

According to Inhelder and Piaget (1958), the ability to represent and manipulate quantitative and numerical *proportions*, which depends on learning a verbally mediated system of numbers and requires computational and logical strategies, emerges only in adolescence, with the acquisition of operational structures (see also Carey, 2009; Feigenson, Dehaene, & Spelke, 2004; Offenbach, Gruen, & Caskey, 1984; Siegler, 1976). To test this hypothesis, Falk and Wilkening (1998) presented 6- to 14-year-old children with a probability-adjustment task. Children were shown an urn of winning and losing beads and were asked to fill in a second urn with beads of one type (i.e., losing or winning beads) to match the beads' proportions in the first urn. The authors found that only at 13 years of age could children proportionally integrate the numbers of winning and losing beads. Indeed, even 15-year-olds very often fail the classic proportional reasoning tasks, in which participants are presented with two boxes containing coloured marbles in different proportions (e.g., Box A contains one white and two black marbles; Box B contains two white and five black marbles) and have to decide from which box to draw, with eyes covered, if they want a white marble (Martignon & Krauss, 2009).

However, evidence of proportional reasoning has been found already in preschoolers (see Goldberg, 1966; Yost, Siegel, & Andrews, 1962), often in modified versions of proportional reasoning problems (e.g., Ginsburg & Rapoport, 1967; Sophian, 2000; Sophian & Wood, 1997) that, for example, framed the problem in terms of analogy (Farrington-Flint, Canobi, Wood, & Faulkner, 2007) or asked children to provide a satisfaction judgement about a gamble rather than to predict a gamble's outcome (Acredolo, O'Connor, Banks, & Horobin, 1989; Schlottmann, 2001). Denison and Xu (2014) showed that even infants have a rudimentary ability to reason about proportions. They presented 10- to 12-month-old infants with a version of the marble problem described above: Infants saw two jars containing different proportions of pink lollipops, for which infants consistently showed preference, and black lollipops. The jars were then covered, and one lollipop was randomly and invisibly removed from each jar and hidden in a separate location. The authors found that infants correctly inferred which jar was more likely to yield the preferred pink lollipop on a single, random draw, based on proportions, and searched in the correct location.

A matter of representation

Previous studies have compared adults' abilities to make qualitative versus quantitative proportional judgements, with mixed results. For example, engineers performed more accurately when a task was presented numerically and they were encouraged to work out explicit equations for making their judgements, compared to when the task was presented visually and without numbers (Hammond, Harnrn, Grassi, & Pearson, 1987; see also Ahl, Moore, & Dixon, 1992). However, Erev, Bornstein, and Wallsten (1993) found that participants' choices between gambles were less optimal when making numerical assessments of probability than when making choices without numbers.

Whether adults' performance in a numerical task is superior to performance in a task not relying on symbolic number representations depends partly on the criteria used to assess performance and partly on other task variables. In particular, one variable that has been shown to be crucial is *numeracy*, that is, participants' general ability to understand and reason with numerical information. People with low numeracy have difficulties using numbers and processing elementary probability expressions (Kutner, Greenberg, Jin, & Paulsen, 2006; Lipkus, Samsa, & Rimer, 2001; Peters *et al.*, 2006; Schwartz, Woloshin, Black, & Welch, 1997; Tuijnman, 2000). Visual and more intuitive displays have been proposed as a potentially promising method for communicating medical risks to such low-numeracy people (Ancker, Senathirajah, Kikafka, & Starren, 2006; Galesic, Garcia-Retamero, & Gigerenzer, 2009; Lipkus & Hollands, 1999; Montori & Rothman, 2005; Reyna, Nelson, Han, & Dieckmann, 2009). Among the different visual displays, icon arrays are gaining in popularity both in medical practice and in public media (Ancker *et al.*, 2006; Edwards, Elwyn, & Mulley, 2002; Elmore & Gigerenzer, 2005; Paling, 2003). Icon arrays are graphical representations consisting of a number of stick figures, faces, circles, or other icons, arranged in rows and columns, symbolizing individuals with a certain property. Icon arrays have been shown to be an effective method for eliminating denominator neglect in adults when conveying treatment risk reduction information (Garcia-Retamero & Galesic, 2009; Garcia-Retamero, Galesic, & Gigerenzer, 2010) and to communicate risk effectively (see Hess, Visschers, & Siegrist, 2011). Galesic *et al.* (2009) found that icon arrays are especially useful for those students and older adults whose numeracy is low compared to the average of their groups.

Previous work indicates that children's reasoning abilities vary as a function of the structure of the representations they are given (Boyer, Levine, & Huttenlocher, 2008; Martignon & Krauss, 2009; Multmeier, Gigerenzer, & Wegwarth, 2014; Ruggeri & Feufel, 2015). Do children generally benefit from visual and more intuitive displays? On the one hand, certain representations of statistical concepts, such as tinker cubes – small plastic Lego-like cubes of different colours that can be assembled to form towers that encode information about one individual in a population (Martignon & Krauss, 2009) – and icon arrays (Multmeier *et al.*, 2014), have been shown to facilitate probabilistic reasoning and inferences in children. On the other hand, Boyer *et al.* (2008) showed that 9-year-olds still have difficulty reasoning proportionally when proportions are presented as discrete quantities (e.g., discrete *units* of juice vs. units of water), whereas even 6- and 7-year-olds demonstrate proportional reasoning abilities when presented with continuous quantities (e.g., *amount* of juice and water, not represented as a sum of single units). These findings suggest that children's difficulties with proportional reasoning might stem from their tendency to compare quantities based on the number of elements in the target quantity rather than on the basis of proportional relations (see also Falk, Yudilevich-Assouline, & Elstein, 2012).

Thus, it is an open question whether visual representations such as icon arrays would improve children's performance in a proportional reasoning task compared to numerical representations such as frequencies (e.g., '10 out of 100'; see Zhu & Gigerenzer, 2006). Based on the literature reviewed above, one possibility is that both children and adults would benefit from the icon-array format. Another possibility is that children, generally with lower numeracy than adults, would make more accurate proportional judgments when presented with icon arrays than when presented with numbers, whereas adults, overall, might be more accurate when presented with the numerical format (as in Erev *et al.*, 1993).

Time-limitation effects

In many real-life situations, decisions have to be made under limited time, and it is known that adults adapt to time limitation using faster and simpler (Ben Zur & Brenitz, 1981; Payne, Bettmann, & Johnson, 1988), but not necessarily more effective (Belling, Suss, & Ward, 2015) strategies. Recent work has shown that children are *ecological learners* – they modify their learning strategies to the characteristics of the task at hand (Horn, Ruggeri, & Pachur, 2016; Nelson, Divjak, Gudmundsdottir, Martignon, & Meder, 2014; Ruggeri & Lombrozo, 2015), and they do so already by age 4 (Ruggeri, Sim, & Xu, 2017). The effects of time limitation on the efficiency of children's reasoning are mixed. Davidson (1996) found that time pressure promoted faster, but not more effective, information search in 7- to 10-year-old children – as for adults. However, Musculus, Ruggeri, Raab, and Lobinger (2017) showed that under time constraints, 6- to 13-year-olds tended to generate fewer and better alternative options in a soccer-related option-generation task.

In the study reported here, we investigated how children's proportional reasoning is affected by time limitation. We hypothesized that participants would be more accurate in the unlimited-time condition where they had more time to make the proportional judgement. We also expected that more time available to make the proportional judgement would benefit adults more than children, because adults, when given more time, might be able to implement more sophisticated and accurate computing strategies that children have not yet mastered (e.g., proportions and fractions; see Ben Zur & Brenitz, 1981; Payne *et al.*, 1988). Moreover, we were intrigued by the possibility that icon arrays might be differentially beneficial depending on the time available to make the proportional judgment.

Basic and sophisticated numerical skills

Basic numerical skills (e.g., estimation) are thought to be building blocks for the acquisition of more sophisticated numerical abilities (e.g., proportional reasoning; Meyer, Salimpoor, Wu, Geary, & Menon, 2010; Passolunghi, Mammarella, & Altoè, 2008; Piazza *et al.*, 2010). Among the basic cognitive abilities and numerical skills contributing to math achievement and potentially to proportional reasoning, the approximate number system (ANS) plays a crucial role. The ANS is a cognitive system that allows one to estimate quantities without language or instruction (Dehaene, 1997; Feigenson *et al.*, 2004; Kibbe & Feigenson, 2014; Libertus & Brannon, 2009). Although ANS representations are inherently imprecise, varying across individuals, this imprecision decreases over the course of development (Halberda & Feigenson, 2008; Lipton & Spelke, 2004) and does not level off until well into adulthood (Halberda, Ly, Wilmer, Naiman, & Germine, 2012). The ANS has been shown to support a variety of number-relevant computations. For example, 11-month-old infants recognize ordered relations among dot arrays, dishabituating when sequences of arrays change from numerically ascending to descending or vice versa (Brannon, 2002). Six-month-old infants can detect a common ratio of blue to yellow shapes across arrays containing different absolute quantities, suggesting a rudimentary division computation (McCrink & Wynn, 2007), and can discriminate eight dots from 16 (a 2:1 ratio), but not eight dots from 12 (a 3:2 ratio; Xu, 2003; Xu & Spelke, 2000). Preschoolers can indicate which of two approximate quantities is larger, even when the quantities are presented in different sensory modalities (Barth, La Mont, Lipton, & Spelke, 2005). Children can use the ANS to reason about arithmetic problems using number words and digits prior to receiving mathematical instruction (Barth *et al.*, 2006; Booth & Siegler, 2008; Gilmore, McCarthy, & Spelke, 2007).

By learning to use symbolic number representations (i.e., Arabic numerals and the place-value system), children become able to solve problems of increasing difficulty, in terms of both numerical range and numerical ratio. It is also assumed that the acquisition of symbolic number formats changes the accuracy and the scaling of the underlying representation from logarithmically compressed to linear scaling (Dehaene, 2003). Yet, even later in development, the ANS continues to play an important role in numerical cognition, maintaining a strong link with symbolic mathematics. Individual differences in the precision of the ANS have been shown to correlate, even in adolescence, with differences in formal math ability as assessed by standardized math tests (Feigenson, Libertus, & Halberda, 2013; Halberda, Mazzocco, & Feigenson, 2008; Libertus, Feigenson, & Halberda, 2013a, 2013b; Starr, Libertus, & Brannon, 2013).

The link between such basic and more sophisticated numeracy skills remains a matter of debate. On the one hand, studies on daily-life choices based on numbers (Banks, O'Dea, & Oldfield, 2010; Gerardi, Goette, & Meier, 2013; Peters *et al.*, 2006; Reyna *et al.*, 2009) and studies on proportional reasoning skills (Bonato, Fabbri, Umiltà, & Zorzi, 2007; Schneider & Siegler, 2010) have not explored the link between these high-level decisions and individual performance on basic numerical tasks. On the other hand, studies on basic numeracy skills have not extensively explored the role of these building blocks in solving comparisons of ratio quantities such as proportions, fractions, and percentages (Halberda *et al.*, 2012; Mazzocco, Feigenson, & Halberda, 2011).

We were interested in investigating whether and how participants' general proportional reasoning performances are correlated with basic numerical and cognitive skills. In particular, we analysed the impact of (1) number sense and ANS aptitude, by testing participants on the Panamath task (Halberda & Feigenson, 2008; Halberda *et al.*, 2008, 2012), which measures the accuracy of participants' choices and response times in a quantity-comparison task; and on the dot enumeration task, which measures the capacity for estimating small numbers and is frequently used as a subtest to assess whether children are affected by dyscalculia (Butterworth, 2003); and (2) the information processing and working memory speed, as measured by the digit-symbol task (Wechsler, 1981). We additionally explored whether these basic numerical and cognitive skills would differentially impact proportional reasoning in our task depending on presentation format (icon arrays vs. frequency) and time limitation.

Method

Participants

Participants were 33 children in second grade (24 female; M_{age} = 7.15 years; SD = 0.71), 23 children in fifth grade (13 female; M_{age} = 10.17 years; SD = 0.58) and 27 adults (14 female; M_{age} = 29.93 years; SD = 5.17). The results of nine additional children and three adults were not included in the analysis because they did not complete the task (n = 2) or because of technical difficulties (n = 10). Children were recruited from local schools in Berlin, Germany. Adult participants were recruited through the participant pool of the Max Planck Institute for Human Development in Berlin, Germany.

Design and procedure

The experimental session consisted of four blocks of 11 trials each. On each trial, participants were shown the contents of two bags, Bag 1 and Bag 2, displayed on a computer screen (see Figure 1). Participants were told that each bag contained a different number of tokens, varying randomly between 50 and 100. Some of these tokens were gold and some were black. Participants were asked to press a button on the keyboard corresponding to the bag from which they wanted to *randomly* draw a token. They were told that they had to select the bag from which it would be more likely to randomly draw a gold token, that is, the bag with the *higher proportion* of gold tokens.

The first trial of each block was always the easiest (i.e., 100% gold tokens in one bag and 0% gold tokens in the other bag) and was used to familiarize the participants with the game and the block-specific manipulation. The results of these first trials were not included in the analysis. Across the following 10 trials, we manipulated the trials' difficulty, that is, the proportion of gold and black tokens contained in the bags, as presented in Table 1. The order of the 10 test trials was randomized within each block; we also randomized the assignment of the proportions of the two bags to Bag 1 (presented on the left) and Bag 2 (presented on the right).

The presented trials were either congruent or incongruent (respectively, about 75% and 25% of the trials). Congruent trials could be solved by comparing only the numerators of the ratios and could be solved with a simple heuristic, that is, 'choose the bag with the highest (absolute) number of gold tokens'. To correctly solve incongruent trials, participants would have to take into account both the numerator and the denominator of the two bags, because the bag containing the highest (absolute) number of desired objects

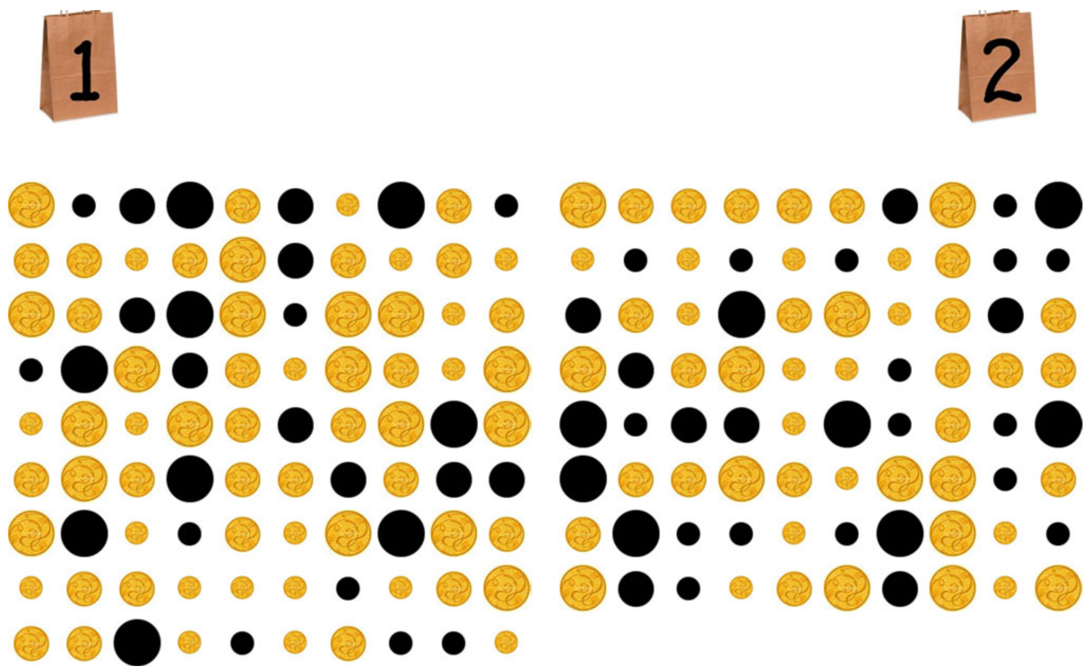

Figure 1. Screenshot of an experimental trial showing an example of a medium–difficult trial in the icon-array condition.

Table 1. Proportions of gold and black tokens contained in the two bags in the 11 trials, displayed by trial difficulty

Difficulty	Difference (%) in gold tokens between Bags 1 and 2	Bag 1			Bag 2		
		Gold tokens (%)	Black tokens (%)	Ratio of bag	Gold tokens (%)	Black tokens (%)	Ratio of bag
Training	100	100	0	–	0	100	–
Easy	80	90	10	9	10	90	9
	60	80	20	4	20	80	4
Medium–easy	40	70	30	2.33	30	70	2.33
	30	80	20	4	50	50	1
Medium	20	80	20	4	60	40	1.50
	20	60	40	1.50	40	60	1.50
Medium–difficult	10	60	40	1.50	70	30	2.33
	10	55	45	1.22	45	55	1.22
Difficult	4	52	48	1.08	48	52	1.08
	2	51	49	1.04	49	51	1.04

would not be the one containing the highest *proportion* (relative amount) of desired objects.

Across the four blocks, we manipulated two independent variables in a 2 $\times$ 2 matrix.

Presentation format

In two blocks, the contents of the bags were presented as *icon arrays* (icon-array condition), that is, a series of pictograms of gold and black tokens arranged in rows and columns (see Figure 1). The arrangement of black and gold tokens on the grid, as well as the size of each pictogram (large, medium, or small), was randomized within trials (for a comparison of performances in randomly vs. systematically arranged icon arrays, see Feldman-Stewart, Brundage, & Zotov, 2007; Feldman-Stewart, Kocovski, McConnell, Brundage, & Mackillop, 2000). In the other two blocks, the contents of the bags were presented in number format as *frequencies*, for example, '90 tokens out of 100 are gold' (frequency condition).

Time limitation

In two blocks, participants could visualize the contents of the bags (displayed either as icon arrays or as frequencies) for as long as they wanted to (unlimited-time condition). Participants could click on the button corresponding to the chosen bag whenever ready. In the other two blocks, the contents of the bags were displayed for only 3 s (limited-time condition). After the 3 s, the contents of the two bags were masked with images representing two bags, between which participants had to choose.

The order of the four blocks (icon array/limited time, icon array/unlimited time, frequency/limited time, and frequency/unlimited time) was pseudorandomized so that we presented the same order to a similar number of participants, to avoid any order effect (see Ahl *et al.*, 1992).

Participants received no immediate feedback on whether they had chosen the 'correct bag', that is, the bag with the higher proportion of gold tokens. Feedback was given at the end of each experimental block, when the number of gold tokens 'drawn' from the selected bags during the 10 test trials was revealed ('Good job! You collected 6 gold tokens!'). Although participants were instructed that a token would be *randomly* drawn from the selected bag, we rewarded participants based on the accuracy of their selection: For each trial, if the bag selected was the one containing a higher proportion of gold tokens, then we told the participants they had drawn a gold token; otherwise, a black token. For each gold token drawn, that is, for each correct selection, participants received one sticker (for children) or 10 cents (for adults). Additionally, adults were given 10 euros to thank them for their participation.

At the end of the experimental session, three independent tests were administered to the participants: (1) the Panamath test (Halberda & Feigenson, 2008; Halberda *et al.*, 2008), (2) the dot enumeration test, and (3) the digit–symbol subtest of the Hamburg-Wechsler-Intelligenztests für Kinder (HAWIK-IV), the German translation and adaptation of the Wechsler Intelligence Scale for Children (Wechsler, 1981).

Panamath test. Panamath is a computerized test that measures participants' number sense and ANS aptitude. Participants have to decide as quickly as possible whether there are more blue dots, displayed on one side (right or left) of the screen, or red dots, displayed on the other side of the screen, across many trials. The test was administered for 8 min, enough time to complete an average of 204 trials (SD = 61). The difficulty level was set at 2 (medium) for all participants. No feedback was given to the participants after the test.

Dot enumeration test. The dot enumeration test is a 36-trial computer task in which participants are asked to compare the number of dots (presented as a visual array of up to nine dots) displayed on half of the screen with the numeral displayed on the other half of the screen. To do so, participants need the capacity to enumerate the set of dots, either by seeing immediately that there are one, two, three, or four dots in the set without counting them (i.e., 'subitizing') or by counting a larger sets of dots. Participants have to press a key as quickly as possible to assess whether the numeral matches the number of dots displayed.

Digit–symbol test. The digit–symbol task presents participants with nine digit–symbol pairs (e.g., '2' and '→'), displayed at the top of a printed sheet, followed by a list of digits. Participants have to write under each digit the corresponding symbol from the pairs shown at the top of the page (e.g., writing under each '2' a '→') proceeding in sequential order (i.e., without skipping any digits), to correctly match as many digit–symbol pairs as possible within 90 s.

Results

For each block, we averaged participants' accuracy results across all trials within the same trial difficulty group (easy to difficult, see Table 1). We then performed a repeated-measures analysis of variance (ANOVA) with the average accuracy as the dependent variable; trial difficulty (five levels: easy, medium–easy, medium, medium–difficult, difficult), time limitation (two levels: limited, unlimited), and presentation format (two levels: icon array, frequency) as within-subject factors; and age group (three levels: younger children, older children, adults) as between-subjects factor.

We found a main effect of age group, $F(2,80) = 31.42$, $p < .001$, $\eta^2 = .44$. A series of Bonferroni-corrected pairwise comparison analyses show that both younger and older children ($M_{younger\_child} = .62$, $SE = .021$; $M_{older\_child} = .67$, $SE = .025$) were overall less accurate than adults ($M_{adults} = .86$, $SE = .023$; $p < .001$). We found no difference between younger and older children ($p = .536$).

Analyses revealed a main effect of trial difficulty, $F(4,320) = 39.42$, $p < .001$, $\eta^2 = .33$. As can be seen in Table 2, participants were less accurate in more difficult trials. We also found an interaction of Trial Difficulty × Age Group, $F(8, 320) = 3.16$, $p = .002$, $\eta^2 = .07$. The age difference was more pronounced in easy to medium–difficult trials (all ps < .001, with decreasing effect size from easy, medium–easy and medium to medium–difficult trials: easy trials: $\eta^2 = .33$; medium–easy trials: $\eta^2 = .39$; medium trials: $\eta^2 = .36$; medium–difficult trials: $\eta^2 = .24$) as compared to difficult trials, $p = .033$, $\eta^2 = .08$.

We found no main effect of presentation format ($p = .082$), but an interaction of Presentation Format × Age Group, $F(2,80) = 4.64$, $p = .012$, $\eta^2 = .10$. Separate repeated-measures ANOVAs by age group confirmed that younger children performed better in the icon-array ($M_{icon} = .66$, $SE = .028$) compared to the frequency ($M_{frequency} = .57$, $SE = .020$) condition, $F(1,33) = 14.19$, $p = .001$, $\eta^2 = .30$; older children did not perform better in one condition over the other ($M_{icon} = .68$, $SE = .037$; $M_{frequency} = .62$, $SE = .043$; $p = .274$), and adults performed slightly better in the frequency ($M_{frequency} = .87$, $SE = .025$) compared to the icon-array ($M_{icon} = .81$, $SE = .021$) condition, $F(1,26) = 2.98$, $p = .096$, $\eta^2 = .10$. We also found an interaction of Presentation Format × Trial Difficulty, $F(4,320) = 3.37$, $p = .010$, $\eta^2 = .04$: In the easier trials, participants performed better in the icon-array compared to the frequency condition, and in the medium–difficult trials, presentation format did not affect

Table 2. Average correct choices made by younger children (7-year-olds), older children (10-year-olds), and adults, displayed as a function of time limitation, presentation format, and trial difficulty

		Age group	Easy trials Mean	Easy trials *SD*	Medium–easy trials Mean	Medium–easy trials *SD*	Medium trials Mean	Medium trials *SD*	Medium–difficult trials Mean	Medium–difficult trials *SD*	Difficult trials Mean	Difficult trials *SD*
Unlimited time	Icon array	Younger children	0.82	0.33	0.70	0.39	0.76	0.36	0.70	0.30	0.56	0.37
		Older children	0.83	0.36	0.78	0.33	0.83	0.29	0.57	0.38	0.57	0.31
		Adults	1.00	0.00	1.00	0.00	0.93	0.18	0.78	0.35	0.54	0.44
	Frequency	Younger children	0.68	0.35	0.67	0.37	0.47	0.41	0.61	0.32	0.56	0.32
		Older children	0.72	0.36	0.74	0.33	0.70	0.36	0.59	0.36	0.61	0.37
		Adults	1.00	0.00	0.96	0.13	0.89	0.21	0.87	0.30	0.87	0.30
Limited time	Icon array	Younger children	0.77	0.33	0.58	0.36	0.68	0.33	0.53	0.39	0.58	0.28
		Older children	0.87	0.27	0.80	0.29	0.67	0.36	0.59	0.42	0.54	0.37
		Adults	1.00	0.00	0.98	0.10	0.85	0.23	0.72	0.32	0.57	0.38
	Frequency	Younger children	0.59	0.34	0.61	0.35	0.55	0.32	0.59	0.34	0.50	0.40
		Older children	0.80	0.29	0.65	0.44	0.59	0.36	0.52	0.38	0.41	0.44
		Adults	1.00	0.00	0.94	0.16	0.91	0.20	0.74	0.38	0.69	0.34

participants' performance, whereas in the difficult trials participants performed slightly better in the frequency condition (see Table 2).

We also found a main effect of time limitation, $F(1,80) = 9.20$, $p = .003$, $\eta^2 = .10$. Participants were more accurate in the unlimited-time ($M_{\text{unlimited}} = .74$, $SE = .016$) compared to the limited-time ($M_{\text{limited}} = .69$, $SE = .015$) condition.

Because, as mentioned above, only 25% of the trials were incongruent, and their distribution across the independent variables was uneven (e.g., there were no incongruent trials among the easy trials), we further analysed the data with a generalized estimating equations model analysis that allowed us to consider all the data available by looking at each trial separately (binary coded as 'correct' or 'incorrect'). Besides the effects already highlighted by the previous analysis, the model revealed a main effect of congruency, $B = 2.72$, Wald $\chi^2 = 223.00$, $p < .001$, indicating that participants were less likely to be accurate in the incongruent trials (correct in 46% of the trials) than in the congruent trials (correct in 79% of the trials). The interaction between congruency and age group, $B = .987$, Wald $\chi^2 = 33.53$, $p < .001$, revealed that incongruence impaired some age groups more than others. In particular, older children and adults were the most affected: the accuracy of older children dropped from 75% in the congruent trials to 37% in the incongruent trials, $\chi^2 = 104.18$, $df = 1$, $p < .001$, and the accuracy of adults from 94% in the congruent trials to 57% in the incongruent trials, $\chi^2 = 211.06$, $df = 1$, $p < .001$. The performance of younger children dropped from 68% in the congruent trials to 40% in the incongruent trials, $\chi^2 = 67.71$, $df = 1$, $p < .001$. Separate one-sample *t*-tests confirmed that adults' performance was at chance level in incongruent trials, $t(26) = 1.43$, $p = .165$, whereas that of younger, $t(33) = -2.20$, $p = .035$, and older, children $t(22) = 3.27$, $p = .003$, was below chance level.

Additional tests

Results of the additional tests can be found in Table 3. We ran separate univariate ANOVAs with performances in the Panamath, dot enumeration, and digit–symbol tests as dependent variables, and age group (three levels: younger children, older children, and adults) as between-subjects factor.

Panamath

We found a significant effect of age group on the Weber fraction (w) score of the Panamath test, $F(2,83) = 6.11$, $p = .003$, $\eta^2 = .13$. A series of Bonferroni-corrected pairwise comparisons showed no significant differences between younger and older children ($p = .31$), or between older children and adults ($p = .33$), but younger children's performances resulted in a higher Weber fraction (indication of a less accurate performance) than adults', $p = .002$.

Dot enumeration

We did not find a significant effect of age group on the average number of correct trials, $p = .480$, possibly due to a ceiling effect: Younger children, older children, and adults all achieved a comparable accuracy of around 32 correct trials out of 36. However, we found a significant age group effect on the average time needed to make the correct decisions on the whole range of numbers, which includes the subitizing range (1–5), $F(2,83) = 34.55$, $p < .001$, $\eta^2 = .46$. A Bonferroni-corrected pairwise comparison showed that younger children needed more time to make the correct decisions than older children ($p < .001$) and that older children, in turn, needed more time to make the correct decisions than adults ($p = .005$).

Digit–symbol

We found a significant effect of age group on participants' performance in the digit–symbol test, $F(2,83) = 73.53$, $p < .001$, $\eta^2 = .65$. A Bonferroni-corrected pairwise comparison showed that younger children obtained a lower score than older children ($p = .053$) and that older children, in turn, obtained a lower score than adults ($p < .001$). We also found a correlation between the results of the Panamath, dot enumeration and digit–symbol tests (see Table 3).

Correlations between the proportional reasoning task and the additional tests

The performance in our main experiment significantly correlated with the performance in all the additional tests. Because we found no differences between the two presentation formats (icon array vs. frequency), we only report the correlation results with the overall performance across conditions (see Table 3). As expected, the percentage of correct trials in our proportional task correlated negatively with the Panamath's Weber fraction across

Table 3. Participants' performance in the proportional reasoning task and on the additional tests (Panamath, dot enumeration and digit–symbol), displayed by age group, and correlations across participants between tests

Test	Younger children		Older children		Adults		Correlations between tests			
	Mean	*SD*	Mean	*SD*	Mean	*SD*	1	2	3	4
1. Our study: average percentage correct trials (%)	61	33	65	35	84	22				
2. Panamath: average Weber fraction	0.35	0.25	0.27	0.14	0.18	0.06	−.486[b]			
3. Dot enumeration: average number of correct trials	31.88	4.46	32.42	3.24	33.04	2.58	.260[a]	−.445[b]		
4. Dot enumeration: average reaction time (ms)	3,082.24	1,009.89	2,176.08	481.57	1,473.47	525.33	−.479[b]	.217[a]	.119	
5. Digit–symbol: average number of completed items	40.12	12.14	47.54	7.37	74.89	53.26	.582[b]	−.365[b]	.159	−.611[b]

Notes. [a]Correlation is significant at the 0.05 level (two-tailed).
[b]Correlation is significant at the 0.01 level (two-tailed).

participants ($p < .001$), indicating an improvement of performance over age that coincided with an increasing number acuity as measured by the Panamath test. The percentage of correct trials in the proportional reasoning task was negatively correlated with the Weber fraction within groups (adults: Pearson's $r = -.027$, $p = .896$; older children: Pearson's $r = -.261$, $p = .218$), but the correlation was significant only for younger children (Pearson's $r = -.488$, $p = .004$).

We found a positive correlation between performance in the main proportional reasoning task and the average number of correct trials in the dot enumeration test ($p = .018$), as well as a negative correlation with reaction time in the dot enumeration test ($p < .001$), significant only across groups.

Similarly, the performance in the proportional reasoning task correlated with higher working memory as measured by the number of completed items in the digit–symbol test ($p < .001$). Again, this correlation was significant across groups only. None of the tests differentially correlated with participants' performance in the four conditions.

Discussion

We investigated the effects of two context variables, presentation format, and time limitation, on the proportional reasoning abilities of 7- and 10-year-old children and adults. In our experimental task, participants had to choose the set of tokens that offered the highest likelihood of drawing a gold token, that is, the set with the greater proportion of gold tokens. We also explored whether and how participants' proportional reasoning skills were correlated with more general and basic numerical and cognitive skills, such as their ANS aptitude and working memory.

As expected, we found a general improvement in accuracy across development. The age difference held across all levels of difficulty but was more pronounced in easy to medium–difficult trials as compared to difficult trials, where even adults' performances dropped dramatically.

We found that in our task, only younger children performed better in the icon-array condition, whereas older children and adults were similarly accurate in both conditions, with adults performing slightly better in the frequency condition. These results are in line with previous eye-tracking and behavioural studies suggesting that in the first intuitive phase of numeric information processing, adults with higher numeracy rely more on numerical information, whereas adults with lower numeracy privilege icon arrays and generally seem to be confused by numbers (Hess *et al.*, 2011; Keller, Kreuzmair, Leins-Hess, & Siegrist, 2014).

On the one hand, the mathematical computing strategies adults and older children might implement are explicit cognitive operations that can only be applied to proportions expressed as numbers and frequencies (see Meert, Grégoire, Seron, & Noël, 2012). In icon arrays, the calculation has to be necessarily more intuitive, unless participants count the icons one by one, which was very challenging in our task because of the larger numbers of icons in the arrays, their variation in size, and their random arrangement. Note that previous literature shows that adults typically perform worse on randomly arranged icon arrays as compared to sorted arrays, where the icons are grouped by type and homogeneous in size (see Feldman-Stewart *et al.*, 2000, 2007).

On the other hand, children might have had trouble understanding frequencies (see Meert, Grégoire, Seron, & Noël, 2013), even those referring to the easiest ratios, just because they were presented in a numerical format. The more intuitive icon-array condition might have helped them to be accurate, at least in the easier trials. Indeed, we found that icon arrays were particularly beneficial in the easier trials. This can be interpreted in two ways: First, the intuitive perceptual computation facilitated by icon arrays might not be fine-grained enough to allow more difficult trials to be computed, because in those trials the difference between proportions is extremely small and therefore difficult to detect when not expressed in a more accurate, numerical format. Second, more difficult trials were more likely to be incongruent: it may be that icon arrays only facilitate performance in congruent trials, those that can be solved just by comparing the proportions' numerators. Generally, sorted and size-homogeneous icon arrays might have further improved younger children's perceptual and intuitive proportional comparison strategies, as well as facilitated adults' more explicit strategies conditional on counting the icons in the arrays.

Not surprisingly, we found that participants were more accurate in the congruent trials, that is, the trials that could be solved by comparing only the first terms of the ratio and with a simple heuristic, that is, 'choose the alternative with the highest (absolute) number of gold tokens'. Our results seem to suggest that the performances of older children and adults were hindered by incongruency more than younger children's. However, this result is likely due to a floor effect: because the younger children's performances were already quite poor in the congruent trials, they could not get much worse in the incongruent trials. Crucially, adults' performance on the incongruent trials was at chance level, suggesting that they did pay attention to proportions after all but failed to compute them correctly. In contrast, in line with results from previous literature (see Falk & Wilkening, 1998; Girotto, Fontanari, Gonzalez, Vallortigara, & Blaye, 2016; Ni & Zhou, 2005), children's performance was *below* chance level, indicating that they might have paid attention only to the numerator. Unfortunately, as our task was not designed to investigate specifically the effect of congruency (e.g., only a small proportion of trials were incongruent, and they were not evenly distributed across the different difficulty levels), the effect of congruency in the data we have available is confounded (see Fontanari, Gonzalez, Vallortigara, & Girotto, 2014). Future research should address this limitation, in order to be able to further investigate the differential benefit of icon arrays for trials of different difficulty.

Finally, we found the hypothesized effect of time limitation, with participants of all age groups being overall more accurate when they were given more time to make the proportional judgement. Interestingly, we found that having more time available did not benefit adults more than children, probably because adults' performance was already pretty accurate in the limited-time condition. We also did not find any interaction of presentation format and time limitation. Generally, our results seem to suggest that adults are not more likely than children to implement more sophisticated and precise proportional reasoning strategies when they have more time available or when stimuli are presented in a frequency versus icon-array format.

Interestingly, we found that the results of our study are correlated with the results of all the other cognitive tests participants performed. Evidently, the abilities necessary to perform well in our proportional reasoning task include a working memory component (measured by the digit–symbol task) and build on participants' number sense (measured by the Panamath task), as well as on the ability to subitize (captured by the dot enumeration task). However, none of these tests is able to explain individual differences in performances in our task *within* age groups (with the exception of the Panamath's Weber fraction, which correlates with the individual performances of younger children).

In sum, we found that both presentation format and time limitation affected children's and adults' proportional reasoning performance. In particular, we found that younger children generally benefited from a visual, intuitive representation of proportions such as icon arrays, whereas adults did better with numbers. Future research should explore the potential benefits of visual representation of proportions and numbers with toddlers and preschoolers and develop a more controlled procedure to test the ability to compare visual representations of ratios, for example, substituting ordered icon arrays (i.e., arranged in rows and columns) with clusters of dots, as in the Panamath task. This would allow better control for factors such as spread and object size and would offer more accurate measures of performance. Overall, our results further encourage research exploring visual formats and interventions to support children's learning and understanding of proportions.

Acknowledgements

We would like to thank Gregor Caregnato, Katherine McMahon, Anselm Rothe, and Anja Blöhm for helping us run the experiments, as well as Anita Todd and Nicolai Bodemer for their support and helpful comments. This research was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme to A.R.

References

Acredolo, C., O'Connor, J., Banks, L., & Horobin, K. (1989). Children's ability to make probability estimates: Skills revealed through application of Anderson's functional measurement methodology. *Child Development*, *60*, 933–945. https://doi.org/10.2307/1131034

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

Ancker, J. S., Senathirajah, Y., Kikafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, *13*, 608–618. https://doi.org/10.1197/jamia.M2115

Banks, J., O'Dea, C., & Oldfield, Z. (2010). Cognitive function, numeracy and retirement saving trajectories. *The Economic Journal*, *120*, F381–F410. https://doi.org/10.1111/j.1468-0297.2010.02395.x

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*, 199–222. https://doi.org/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 Science of the USA*, *102*, 14116–14121. https://doi.org/10.1073/pnas.0505512102

Belling, P. K., Suss, J., & Ward, P. (2015). Advancing theory and application of cognitive research in sport: Using representative tasks to explain and predict skilled anticipation, decision-making, and option-generation behavior. *Psychology of Sport and Exercise*, *16*, 45–59. https://doi.org/10.1016/j.psychsport.2014.08.001

Ben Zur, H., & Brenitz, S. J. (1981). The effects of time pressure on risky choice behavior. *Acta Psychologica*, *47*, 89–104. https://doi.org/10.1016/0001-6918(81)90001-9

Betsch, T., & Lang, A. (2013). Utilization of probabilistic cues in the presence of irrelevant information: A comparison of risky choice in children and adults. *Journal of Experimental Child Psychology*, *115*, 108–125. https://doi.org/10.1016/j.jecp.2012.11.003

Betsch, T., Lang, A., Lehmann, A., & Axmann, J. M. (2014). Utilizing probabilities as decision weights in closed and open information boards: A comparison of children and adults. *Acta Psychologica*, *153*, 74–86. https://doi.org/10.1016/j.actpsy.2014.09.008

Bonato, M., Fabbri, S., Umiltà, C., & Zorzi, M. (2007). The mental representation of numerical fractions: Real or integer? *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1410–1419. https://doi.org/10.1037/0096-1523.33.6.1410

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*, 1016–1031. https://doi.org/10.1111/j.1467-8624.2008.01173.x

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

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

Butterworth, B. (2003). *Dyscalculia screener*. London, UK: nFER-Nelson.

Carey, S. (2009). *The origin of concepts*. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Davids, S. L., Schapira, M. M., McAuliffe, T. L., & Nattinger, A. B. (2004). Predictors of pessimistic breast cancer risk perceptions in a primary care population. *Journal of General Internal Medicine*, *19*, 310–315. https://doi.org/10.1111/j.1525-1497.2004.20801.x

Davidson, D. (1996). The effects of decision characteristics on children's selective search of predecisional information. *Acta Psychologica*, *92*(3), 263–281. http://doi.org/10.1016/0001-6918(95)00014-3

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, *7*, 145–147. https://doi.org/10.1016/S1364-6613(03)00055-X

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

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

Denison, S., & Xu, F. (2010a). Integrating physical constraints in statistical inference by 11-month-old infants. *Cognitive Science*, *34*, 885–908. https://doi.org/10.1111/j.1551-6709.2010.01111.x

Denison, S., & Xu, F. (2010b). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, *13*, 798–803. https://doi.org/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. https://doi.org/10.1016/j.cognition.2013.12.001

Eaves, B., & Shafto, P. (2012). Unifying pedagogical reasoning and epistemic trust. In F. Xu & T. Kushnir (Eds.). *Rational constructivism in cognitive development. Advances in child development and behavior* (Vol. 43, pp. 295–320). Waltham, MA: Academic Press. https://doi.org/10.1016/B978-0-12-397919-3.00011-3

Edwards, A., Elwyn, G., & Mulley, A. (2002). Explaining risks: Turning numerical data into meaningful pictures. *British Medical Journal*, *324*, 827–830. https://doi.org/10.1136/bmj.324.7341.827

Elmore, J. G., & Gigerenzer, G. (2005). Benign breast disease – The risks of communicating risk. *New England Journal of Medicine*, *353*, 297–299. https://doi.org/10.1056/NEJMe058111

Erev, I., Bornstein, G., & Wallsten, T. S. (1993). The negative effect of probability assessments on decision quality. *Organizational Behavior and Human Decision Processes*, *55*, 78–94. https://doi.org/10.1006/obhd.1993.1025

Falk, R., & Wilkening, F. (1998). Children's construction of fair chances: Adjusting probabilities. *Developmental Psychology*, *34*, 1340–1357. https://doi.org/10.1037/0012-1649.34.6.1340

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*, 207–233. https://doi.org/10.1007/s10649-012-9402-1

Farrington-Flint, L., Canobi, K. H., Wood, C., & Faulkner, D. (2007). The role of relational reasoning in children's addition concepts. *British Journal of Developmental Psychology*, *25*, 227–246. https://doi.org/10.1348/026151006X108406

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

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives*, *7*, 74–79. https://doi.org/10.1111/cdep.12019

Feldman-Stewart, D., Brundage, M. D., & Zotov, V. (2007). Further insight into the perception of quantitative information: Judgments of gist in treatment decisions. *Medical Decision Making*, *27*, 34–43. https://doi.org/10.1177/0272989X06297101

Feldman-Stewart, D., Kocovski, N., McConnell, B. A., Brundage, M. D., & Mackillop, W. J. (2000). Perception of quantitative information for treatment decisions. *Medical Decision Making*, *20*, 228–238. https://doi.org/10.1177/0272989X0002000208

Fontanari, L., Gonzalez, M., Vallortigara, G., & Girotto, V. (2014). Probabilistic cognition in two indigenous Mayan groups. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 17075–17080. https://doi.org/10.1073/pnas.1410583111

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, *28*, 210–216. https://doi.org/10.1037/a0014474

Garcia-Retamero, R., & Galesic, M. (2009). Communicating treatment risk reduction to people with low numeracy skills: A cross-cultural comparison. *American Journal of Public Health*, *99*, 2196–2202. https://doi.org/10.2105/AJPH.2009.160234

Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2010). Do icon arrays help reduce denominator neglect?. *Medical Decision Making*, *30*, 672–684. https://doi.org/10.1177/0272989X10369000

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 11267–11271. https://doi.org/10.1073/pnas.1220568110

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature*, *447*, 589–591. https://doi.org/10.1038/nature05850

Ginsburg, H., & Rapoport, A. (1967). Children's estimate of proportions. *Child Development*, *38*, 205–212. https://doi.org/10.2307/1127141

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. https://doi.org/10.1016/j.cognition.2016.03.010

Girotto, V., & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition*, *106*(1), 325–344. https://doi.org/10.1016/j.cognition.2007.02.005

Goldberg, S. (1966). Probability judgments by preschool children: Task conditions and performance. *Child Development*, *37*, 151–168. https://doi.org/10.2307/1126436

Gonzalez, M., & Girotto, V. (2011). Combinatorics and probability: Six- to ten-year-olds reliably predict whether a relation will occur. *Cognition*, *120*, 372–379. https://doi.org/10.1016/j.cognition.2010.10.006

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085–1108. https://doi.org/10.1037/a0028044

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

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. https://doi.org/10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 11116–11120. https://doi.org/10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668. https://doi.org/10.1038/nature07246

Hammond, K. R., Harnrn, R. M., Grassi, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, *17*, 753–770. https://doi.org/10.1109/TSMC.1987.6499282

Hess, R., Visschers, V. H. M., & Siegrist, M. (2011). Risk communication with pictographs: The role of numeracy and graph processing. *Judgment and Decision Making*, *6*, 263–274.

Hibbard, J. H., Peters, E. M., Slovic, P., & Finucane, M. L. (2001). Making health care report cards easier to use. *The Joint Commission Journal on Quality Improvement*, *27*, 591–604. https://doi.org/10.1016/S1070-3241(01)27051-5

Horn, S. S., Ruggeri, A., & Pachur, T. (2016). The development of adaptive decision making: Recognition-based inference in children and adolescents. *Developmental Psychology*, *52*, 1470–1485. https://doi.org/10.1037/dev0000181

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books. https://doi.org/10.1037/10034-000

Keller, C., Kreuzmair, C., Leins-Hess, R., & Siegrist, M. (2014). Numeric and graphic risk information processing of high and low numerates in the intuitive and deliberative decision modes: An eye-tracker study. *Judgment and Decision Making*, *9*, 420.

Kibbe, M. M., & Feigenson, L. (2014). Young children "solve for x" using the approximate number system. *Developmental Science*, *18*, 38–49. https://doi.org/10.1111/desc.12177

Kushnir, T., & Gopnik, A. (2005). Children infer causal strength from probabilities and interventions. *Psychological Science*, *16*, 678–683. https://doi.org/10.1111/j.1467-9280.2005.01595.x

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

Kutner, M., Greenberg, E., Jin, Y., & Paulsen, C. (2006). *The health literacy of America's adults: Results from the 2003 national assessment of adult literacy* (NCES Report No. 2006–483). Washington, DC: Government Printing Office.

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

Libertus, M. E., Feigenson, L., & Halberda, J. (2013a). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences*, *25*, 126–133. https://doi.org/10.1016/j.lindif.2013.02.001

Libertus, M. E., Feigenson, L., & Halberda, J. (2013b). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology*, *116*, 829–838. https://doi.org/10.1016/j.jecp.2013.08.003

Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *Journal of the National Cancer Institute. Monographs*, *25*, 149–163. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024191

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37–44. https://doi.org/10.1177/0272989X0102100105

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

Martignon, L., & Krauss, S. (2009). Hands-on activities for fourth graders: A tool box for decision-making and reckoning with risk. *International Electronic Journal of Mathematics Education*, *4*, 227–258.

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, *82*, 1224–1237. https://doi.org/10.1111/j.1467-8624.2011.01608.x

McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science*, *18*, 740–745. https://doi.org/10.1111/j.1467-9280.2007.01969.x

Meert, G., Grégoire, J., Seron, X., & Noël, M.-P. (2012). The mental representation of the magnitude of symbolic and nonsymbolic ratios in adults. *Quarterly Journal of Experimental Psychology*, *65*, 702–724. https://doi.org/10.1080/17470218.2011.632485

Meert, G., Grégoire, J., Seron, X., & Noël, M. P. (2013). The processing of symbolic and nonsymbolic ratios in school-age children. *PLoS One*, *8*, e82002. https://doi.org/10.1371/journal.pone.0082002

Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learning and Individual Differences*, *20*, 101–109. https://doi.org/10.1016/j.lindif.2009.08.004

Montori, V. M., & Rothman, R. L. (2005). Weakness in numbers – The challenge of numeracy in healthcare. *Journal of General Internal Medicine*, *11*, 1071–1072. https://doi.org/10.1111/j.1525-1497.2005.051498.x

Multmeier, J., Gigerenzer, G., & Wegwarth, O. (2014). *Do icon arrays improve Bayesian reasoning in children? Towards ecological design*, Manuscript in preparation.

Musculus, L., Ruggeri, A., Raab, M., & Lobinger, B. (2017). *A developmental perspective on option generation and selection: Children conform to the predictions of the take-the-first heuristic*, Manuscript submitted for publication.

Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., & Meder, B. (2014). Children's sequential information search is sensitive to environmental probabilities. *Cognition*, *130*, 74–80. https://doi.org/10.1016/j.cognition.2013.09.007

Ni, Y., & Zhou, Y. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, *40*, 27–52. https://doi.org/10.1207/s15326985ep4001_3

Offenbach, S. I., Gruen, G. E., & Caskey, B. J. (1984). Development of proportional response strategies. *Child Development*, *55*, 963–972. https://doi.org/10.2307/1130148

Okan, Y., Garcia-Retamero, R., Galesic, M., & Cokely, E. (2012). When higher bars are not larger quantities: On individual differences in the use of spatial information in graph comprehension. *Spatial Cognition and Computation: An Interdisciplinary Journal*, *12*, 195–218. https://doi.org/10.1080/13875868.2012.659302

Paling, J. (2003). Strategies to help patients understand risks. *British Medical Journal*, *327*, 745–748. https://doi.org/10.1136/bmj.327.7417.745

Passolunghi, M. C., Mammarella, I. C., & Altoè, G. (2008). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Developmental Neuropsychology*, *33*, 229–250. https://doi.org/10.1080/87565640801982320

Payne, J. W., Bettmann, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*, 407–413. https://doi.org/10.1111/j.1467-9280.2006.01720.x

Piazza, M., Facoetti, A., Trussardi, A., Berteletti, I., Conte, S., Lucangeli, D., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*, 33–41. https://doi.org/10.1016/j.cognition.2010.03.012

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*, 943–973. https://doi.org/10.1037/a0017327

Ruggeri, A., & Feufel, M. A. (2015). How basic-level objects facilitate question-asking in a categorization task. *Frontiers in Psychology*, *6*, 918. https://doi.org/10.3389/fpsyg.2015.00918

Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143*, 203–216. https://doi.org/10.1016/j.cognition.2015.07.004

Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology*, *52*, 2159–2173. https://doi.org/10.1037/dev0000240

Ruggeri, A., Sim, Z. L., & Xu, F. (2017). "Why is Toma late to school again?" Preschoolers identify the most informative questions. *Developmental Psychology*, *53*, 1620–1632. https://doi.org/10.1037/dev0000340

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

Schneider, M., & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1227–1238. https://doi.org/10.1037/a0018170

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, *43*, 1124–1139. https://doi.org/10.1037/0012-1649.43.5.1124

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*, 966–972. https://doi.org/10.7326/0003-4819-127-11-199712010-00003

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*, 481–520. https://doi.org/10.1016/0010-0285(76)90016-5

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

Sophian, C., & Wood, A. (1997). Proportional reasoning in young children: The parts and the whole of it. *Journal of Educational Psychology*, *89*, 309–317. https://doi.org/10.1037/0022-0663.89.2.309

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 18116–18120. https://doi.org/10.1073/pnas.1302751110

Teglas, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, *332*, 1054–1058. https://doi.org/10.1126/science.1196404

Tuijnman, A. (2000). *Benchmarking adult literacy in America: An international comparative study*. Washington, DC: International Adult Literacy Survey, Department of Education, OVAE.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale – Revised: Manual*. New York, NY: Psychological Corporation.

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, *89*, B15–B25. https://doi.org/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*, 97–104. https://doi.org/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. https://doi.org/10.1073/pnas.0704450105

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

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

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

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

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, *98*, 287–308. https://doi.org/10.1016/j.cognition.2004.12.003

Received 2 December 2016; revised version received 9 December 2017