Training nonsymbolic proportional reasoning in children and its effects on their symbolic math abilities

Camilo Gouet\textsuperscript{a,\textdagger}, Salvador Carvajal\textsuperscript{b}, Justin Halberda\textsuperscript{b}, Marcela Peña\textsuperscript{a}

\textsuperscript{a} Laboratorio de Neurociencias Cognitivas, Escuela de Psicología, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Macul, Santiago, Chile
\textsuperscript{b} Department of Psychological and Brain Sciences, The Johns Hopkins University, Baltimore, MD 21218, USA

A R T I C L E   I N F O

Keywords:
Math cognition
Training
Proportions
Fractions
Symbols
Children

A B S T R A C T

Our understanding of proportions can be both symbolic, as when doing calculations in school mathematics, or intuitive, as when folding a bed sheet in half. While an understanding of symbolic proportions is crucial for school mathematics, the cognitive foundations of this ability remain unclear. Here we implemented a computerized training game to test a causal link from intuitive (nonsymbolic) to symbolic proportional reasoning and other math abilities in 4th grade children. An experimental group was trained in nonsymbolic proportional reasoning (PR) with continuous extents, and an active control group was trained on a remarkably similar nonsymbolic magnitude comparison. We found that the experimental group improved at nonsymbolic PR across training sessions, showed near transfer to a paper-and-pencil nonsymbolic PR test, transfer to symbolic proportions, and far transfer to geometry. The active control group showed only a predicted far transfer to geometry. In a second experiment, these results were replicated with an independent cohort of children. Overall this study extends previous correlational evidence, suggesting a functional link between nonsymbolic PR on one hand and symbolic PR and geometry on the other.

1. Introduction

Understanding the cognitive origins of math knowledge is an important challenge for cognitive science and education. Converging evidence from cognitive psychology, cross-cultural studies and neuroscience suggest that this knowledge is built in part from a set of primitive systems for quantification, that we share with many animal species and express early on in development (Dehaene & Brannon, 2011; Núñez & Lakoff, 2000). Most studies that have investigated the links between nonsymbolic and symbolic math abilities have focused on tasks involving absolute quantities, such as estimating the number of elements in a collection of dots, or comparing line lengths, and whether the performance on these tasks relates to symbolic math abilities, particularly whole number knowledge (e.g. Feigenson, Dehaene, & Spelke, 2004; for reviews see De Smedt, Noël, Gilmore, & Ansari, 2013; Lyons, Price, Vaessen, Blomert, & Ansari, 2014). Little is known, in contrast, about how representing and manipulating nonsymbolic relative quantities (e.g. a collection with one-half red balls and one-half blue balls) relates to our ability to process these quantities in a symbolic way (e.g. ½ red balls, ½ blue balls), a crucial achievement in school-mathematics (Siegler et al., 2012). Recent correlational studies have started filling this gap, showing positive associations between nonsymbolic and symbolic proportional reasoning in children and adults (Matthews, Lewis, & Hubbard, 2016; Möhring, Newcombe, Levine, & Frick, 2015). Here, we implemented a computerized training program in 4th grade children to extend these observations and to assess a causal link from intuitive to symbolic proportional reasoning. We further investigated if these intuitions for proportions may be causally connected to other math abilities such as arithmetic and geometry.

Symbolic proportional reasoning is typically taught starting from 3rd grade (Siegler et al., 2012). This ability is considered pivotal in school mathematics, representing a capstone of basic school mathematics, and a platform for more advanced topics such as algebra (Lesh, Post, & Behr, 1988; Siegler et al., 2012). As it implies recognizing ordinal and equivalence relations between rational expressions such as fractions, ratios, and rates, symbolic proportional reasoning is intimately associated with the conceptual knowledge of fractions, which in turn may serve to develop the procedural knowledge of fractions (i.e. the ability to compute arithmetic operations) (e.g. Hallett, Nunes, & Bryant, 2010). Despite its tremendous importance for school mathematics, it is well-known that children struggle for years to master these abilities, and even adolescents and adults struggle with them (Siegler, Fazio, Bailey, & Zhou, 2013).

Nonsymbolic proportional reasoning appears to be functional much earlier. For instance, several studies have shown that children and...
adults can reason intuitively with proportions (Hollands & Dyre, 2000; Singer-Freeman & Goswami, 2001). Six- to 7-year-old can recognize proportional relations among geometric figures (Sophian, 2000), can perform a match-to-sample task involving rectangles filled in various proportions (Spinillo & Bryant, 1991) and can succeed in making probability judgments (Jeong, Levine, & Huttenlocher, 2007). Indeed, studies with infants reveal that nonsymbolic proportional abilities may be in place quite early – for instance, by 6-months of age infants can discriminate between two ratios of discrete elements (McCrink & Wynn, 2007), and by 8–12 months they show some probabilistic reasoning which requires representing proportions (Denison & Xu, 2014; Téglás, Girotto, Gonzalez, & Bonatti, 2007).

Although the idea of a relation between nonsymbolic and symbolic proportional reasoning can be traced back at least to Piaget's days (Inhelder & Piaget, 1958), surprisingly only a handful of studies have empirically tested for this relation (Ahl, Moore, & Dixon, 1992; Fazio, Bailey, Thompson, & Siegler, 2014; Matthews et al., 2016; Möhring et al., 2015). Interestingly, all these studies have reported significant positive associations between these two abilities. For instance, the study by Matthews et al. (2016) found that accuracy at comparing nonsymbolic fractions in college students positively correlated with performance in four different symbolic fractions tests, even after accounting for inhibitory control and absolute magnitude discrimination abilities. And, Möhring et al. (2015) showed that a similar relation holds in children of 8–10 years, who are just learning formal symbolic fractions. Specifically, they found that the amount of error in estimating proportional relations between two-colored columns and a horizontal line, negatively correlated with performance in a symbolic fractions test. These correlational studies complement the evidence from neuroscience suggesting that nonsymbolic and symbolic proportions activate similar fronto-parietal regions (see review in Jacob, Vallentin, & Nieder, 2012). Together, these studies open the door to determining if intuitive nonsymbolic proportional abilities are causally linked with their understanding of symbolic proportions.

Thus, the first goal of our study was to test the trainability of nonsymbolic proportional reasoning, and the second goal was to test a causal link from nonsymbolic to symbolic proportional reasoning in 4th grade children (9–10 years). We implemented a computerized training program following recent successful interventions of this kind in children and adults (Hyde et al., 2014; Khaman, Hanif, Spek, Bertelleti, & Hyde, 2016; Park & Brannon, 2013, 2014; Wang et al., 2016). One subject was discarded because of failure to complete the pre-training assessments. The remaining participants were pseudo-randomly allocated (see below) to either one of two training groups: PR-training (29) and MC-training (27). One subject from PR-training group failed to complete the training program and so was removed from analyses. Three subjects from PR-training did not complete some specific cognitive tests, because they were absent from the school during the evaluation days. Specifically: one subject did not complete the post-training ANS task; two subjects did not complete the post-training Nonsymbolic Proportions test (paper and pencil), Symbolic Proportions, Verbal Analogy and the ANS task. One of the latter subjects did not complete the pre-test Symbolic Proportions. These subjects were still included in the sample, for those tests where they had completed the relevant measures. Children and their parents/caregivers provided written consent before participating in the study. At the end of the study children received a small gift for their participation.

2. Method

2.1. Participants

 Fifty-seven healthy, middle socioeconomic status fourth-grade children participated in the study, recruited from a public school in Santiago, Chile (37 females, Mean age = 9.5 years, Range: 9 to 12 years). This sample size fulfilled the requirement of having at least 50 participants (25 per training group), based on previous comparable and successful training interventions in the field both with children and adults (Hyde et al., 2014; Khaman, Hanif, Spek, Bertelleti, & Hyde, 2016; Park & Brannon, 2013, 2014; Wang et al., 2016). Our experimental group completed 5 days of nonsymbolic Proportional Reasoning training and in general from school practice and assignments and then iteratively allocated each child to the experimental or active control group, equating for their pre-training tests’ scores. Statistical analyses showed no differences between groups in pre-test evaluations across all cognitive assessments (see Table S1).

2.2. Study design

 We implemented a computerized training program along with a pre- and post-test design. Before and after the training, children completed a battery of math and verbal tests (see below). Children were pseudo-randomly assigned to either the proportional reasoning group (PR-training) or the magnitude comparison group (MC-training) while ensuring roughly equal scores from the pre-training assessments. First, we ordered children according to their performance in pre-training assignments and then iteratively allocated each child to the experimental or active control group, equating for their pre-training tests’ scores. Statistical analyses showed no differences between groups in pre-test evaluations across all cognitive assessments (see Table S1).

2.3. Tasks and procedures

2.3.1. Training program

2.3.1.1. Proportional reasoning (PR-training). The computerized proportional reasoning task was inspired by the paper-and-pencil task developed by Möhring et al. (2015). The goal was to measure children’s accuracy at transforming two vertical extents (i.e., water and juice) into an estimate of proportional strength (i.e., strength of juice flavor) on a horizontal response bar. In each trial of this task, children saw a column divided into two differently colored sections appear in the middle of a computer screen (Fig. 1). The upper section was colored blue to represent water. The color of the lower section varied according to the type of fruit juice used in the session (e.g. red for strawberry, yellow for pineapple, etc. see below). Below the column there was a black
horizontallinethatchildrenusedtoindicatetheirresponses.Ontheleft
side of the line there was a cartoon showing a cup of water with a small
amount of fruit in it (indicating a lightly-flavored fruit drink); on the
right side, there was a similar cartoon with relatively more fruit
(indicating a strongly-flavored fruit drink). The children's task was to
indicate, by clicking on the line, the relative amount of fruit flavor in
the mixture of juice and water on each trial (Fig. 1). Children controlled
the onset of the stimuli by pressing the space bar of a keyboard located
below the screen. Stimuli remained on the screen until the child
responded with a timeout of 5000 ms.

Following Möhring et al. (2015), trials in the PR-training task varied
two factors: the presented proportion (ranging from 0 to 1) and the
scaling factor (i.e., the size of the column compared to the black re-
response line). For the presented proportions, we used 12 different pro-
portions of juice to the total amount that spanned the possibilities from
nearly 0 (all water) to nearly 1 (all fruit). The presented proportions
were chosen from among the following 12 bins: 0.06, 0.15, 0.22, 0.28,
0.36, 0.45, 0.55, 0.67, 0.72, 0.79, 0.84, 0.93. To increase the variability
of the presented proportions, and to decrease potential reliance on long-
term memory for any particular training values used across training, the
specific proportions presented were chosen with some variability
around these 12 values for each session. We also varied the scaling
factor (inspired by Möhring et al. (2015)), which was defined as the
ratio between the total height of the column and the width of the
ranking bar (i.e., the black response line, Fig. 1). For example, a scaling
factor of 1:4 meant that the total height of the column was a quarter of
the ranking bar's width. We included this factor because it has been
shown that the error of intuitive proportion estimation increases with
the scaling factor (Möhring et al., 2015). The scaling factors used were:
1:1; 1:1.5; 1:4. Thus in total we presented 36 trials per session, com-
bining 12 proportions with 3 scaling factors. For variety, we also varied
the width of the columns in two levels – with either a narrow or wide
column (see Supplemental Material for a Movie of the training tasks and
a full list of stimuli used).

Following Möhring et al. (2015), trials in the PR-training task varied
two factors: the presented proportion (ranging from 0 to 1) and the
scaling factor (i.e., the size of the column compared to the black re-
response line). For the presented proportions, we used 12 different pro-
portions of juice to the total amount that spanned the possibilities from
nearly 0 (all water) to nearly 1 (all fruit). The presented proportions
were chosen from among the following 12 bins: 0.06, 0.15, 0.22, 0.28,
0.36, 0.45, 0.55, 0.67, 0.72, 0.79, 0.84, 0.93. To increase the variability
of the presented proportions, and to decrease potential reliance on long-
term memory for any particular training values used across training, the
specific proportions presented were chosen with some variability
around these 12 values for each session. We also varied the scaling
factor (inspired by Möhring et al. (2015)), which was defined as the
ratio between the total height of the column and the width of the
ranking bar (i.e., the black response line, Fig. 1). For example, a scaling
factor of 1:4 meant that the total height of the column was a quarter of
the ranking bar's width. We included this factor because it has been
shown that the error of intuitive proportion estimation increases with
the scaling factor (Möhring et al., 2015). The scaling factors used were:
1:1; 1:1.5; 1:4. Thus in total we presented 36 trials per session, com-
bining 12 proportions with 3 scaling factors. For variety, we also varied
the width of the columns in two levels – with either a narrow or wide
column (see Supplemental Material for a Movie of the training tasks and
a full list of stimuli used).

Following Möhring et al. (2015), trials in the PR-training task varied
two factors: the presented proportion (ranging from 0 to 1) and the
scaling factor (i.e., the size of the column compared to the black re-
response line). For the presented proportions, we used 12 different pro-
portions of juice to the total amount that spanned the possibilities from
nearly 0 (all water) to nearly 1 (all fruit). The presented proportions
were chosen from among the following 12 bins: 0.06, 0.15, 0.22, 0.28,
0.36, 0.45, 0.55, 0.67, 0.72, 0.79, 0.84, 0.93. To increase the variability
of the presented proportions, and to decrease potential reliance on long-
term memory for any particular training values used across training, the
specific proportions presented were chosen with some variability
around these 12 values for each session. We also varied the scaling
factor (inspired by Möhring et al. (2015)), which was defined as the
ratio between the total height of the column and the width of the
ranking bar (i.e., the black response line, Fig. 1). For example, a scaling
factor of 1:4 meant that the total height of the column was a quarter of
the ranking bar's width. We included this factor because it has been
shown that the error of intuitive proportion estimation increases with
the scaling factor (Möhring et al., 2015). The scaling factors used were:
1:1; 1:1.5; 1:4. Thus in total we presented 36 trials per session, com-
bining 12 proportions with 3 scaling factors. For variety, we also varied
the width of the columns in two levels – with either a narrow or wide
column (see Supplemental Material for a Movie of the training tasks and
a full list of stimuli used).

The stimuli in each session were presented in two blocks of 18 trials,
pseudo-randomly arranged in increasingly difficulty. Thus, each block
started with easier trials (predominantly with scaling 1:1) and moved
towards more difficult ones (predominantly scaling factor 1:4). We
chose this arrangement of trials (easy to hard) because it has been
shown that moving training from easy to hard trials promotes better
performances than pure random arrangements in an intuitive discrete
magnitude comparison task (Odic, Hock, & Halberda, 2014), and fa-
cilitates transfer to a symbolic task (Wang et al., 2016). All children
received the same set of trials across all sessions.

<table>
<thead>
<tr>
<th>Training (5 Days)</th>
<th>Training Prediction</th>
<th>Pretest/Posttest Assessments</th>
<th>Purpose</th>
<th>Prediction PR-Training</th>
<th>Prediction MC-Training</th>
<th>Group Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsymbolic Proportional Reasoning (PR-training)</td>
<td>Improvements in RT and Accuracy throughout training</td>
<td>Nonsymbolic Proportions</td>
<td>PR Near Transfer</td>
<td>Improvement</td>
<td>No Change</td>
<td>Yes</td>
</tr>
<tr>
<td>Nonsymbolic Magnitude Comparison (MC-training)</td>
<td>Improvements in RT and Accuracy throughout training</td>
<td>Geometry</td>
<td>PR &amp; MC Far Transfer</td>
<td>Improvement</td>
<td>Improvement</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Table 1
Training measures and select pre/post training measures and predictions.

C. Gouet, et al.
Cognition 197 (2020) 104154

Fig. 1. Left, sample of children playing the computer training games. Right, temporal structure of a trial in each training task (PR-training, MC-training). The images show samples from the tasks. The stimulus in these examples has a proportion of juice to total amount of 0.4. The scaling factor is 1:1.5. Note that for the proportions task the feedback bar has a green region denoting the locations for a highly precise response and a yellow region denoting a less precise response. For the magnitude comparison task, the feedback bar has a green region on the correct side. In these examples the responses given by the hypothetical subjects are wrong – i.e., outside of the green and yellow regions. A sound was played, according to the correctness of response, along with this visual feedback.
Children received audio-visual feedback after each response. There were three different forms of feedback according to the error of response (distance between the response and the correct position). No matter where the child responded, a rectangle appeared around the correct response on the response bar (Fig. 1), with the region closest to the correct response colored green to indicate a higher accuracy, surrounded by a yellow region on either side indicating a less accurate response. If the response fell in the correct region, an encouraging beep was played; if it fell in the yellow region a pleasant beep was played, and otherwise, an unpleasant beep was played. With each training session the width of the feedback rectangle was progressively narrowed, starting from being 15% of the ranking bar's width in the first session, to being 7.5% of the ranking bar's width in the fifth session. Thus, in each successive session children had to respond more accurately in order to get encouraging or pleasant feedback. To keep children engaged with the task, we also gave them a running summary of their performance during the session: after 9 consecutive trials a bar graph appeared showing three columns whose heights were proportional to the number of green, yellow, and outside responses children had obtained up to that point (no numbers (i.e. digits) were included in the running summary).

For every trial, we recorded the position of the click response in the ranking bar and the reaction time.

### 2.3.1.2. Magnitude comparison (MC-training)
We designed the Magnitude Comparison training (MC-training) to be as similar as possible to the Proportional Reasoning training (PR-training). In each trial of the magnitude comparison training task, children saw a two-colored column appear in the middle of a computer screen, as in the proportions task (Fig. 1). In this task, however, they had to indicate if there was more juice or water in the mixture. Responses were given by clicking on the right or the left end of the ranking bar respectively. We included this bar to increase perceptual and response similarity with the PR-training task. Importantly, upon horizontal mouse displacement, the response indicator moved discontinuously (it “jumped”) from one end of the ranking line to the other end, without passing through intermediate positions (see a sample video in Supplemental Material). This way we prompted the binary nature of the task. As shown in Results, children’s response patterns displayed typical signatures of magnitude Comparison tasks (adhering to Weber’s law), thus confirming that they understood the task. Next to the left end of the ranking line there was a cartoon with a cup of water; next to the right end a similar cartoon with fruits inside the cup (Fig. 1). The onset of the stimulus was controlled by children as in proportion task, and had the same timeout (5000 ms).

In each session, we included stimuli with 5 unique proportions of juice to total amount. There were 4 trials for the easiest proportion and 8 trials for each of the other proportions, totaling 36 trials per session. With each session we presented a new set of proportions of increasing difficulty (approaching to 0.5). Thus, in the first session they ranged from 0.41 to 0.59, while in the fifth session they ranged from 0.448 to 0.552 (see Supplementary Material for a video and the full list of stimuli). To be more consistent with the literature, in the Results section we present ratio values instead of proportions, by dividing the winning color size by the losing color size. To make the stimuli as similar as possible to the PR-training task, we also varied stimuli according to the scaling factorconst and width. We included four scaling factors 1:1; 1:1.5; 1:3; 1:4 and two levels of width. For this task, however, we expected the performance to be primarily influenced by the ratio between the amounts (i.e. Weber’s law), and not by the total amount involved in the comparison. This is a critical difference compared to the PR-training. Although the overall similarity of the MC-training to the PR-training, the differential successful transfer to either Nonsymbolic (near transfer) or Symbolic Proportional Reasoning (transfer) will demonstrate noteworthy specificity for PR-training.

The order of trials was similar to that in PR-training: two blocks of 18 trials, pseudo-randomly arranged in increasing difficulty. Children received audio-visual feedback after each response. A small green rectangle appeared on the correct side after the response with no extended yellow region (Fig. 1). If the response was correct, an encouraging beep was played; if it was incorrect, an unpleasant beep was played. As in the PR-training task, children received a running summary of their performance every 9 trials, in this case with two columns indicating the accumulated number of correct and incorrect responses (plus a small fixed column in the middle to increase perceptual similarity with the running summary from PR-training). For every trial, we recorded the side of their responses (left or right) and the reaction time.

### 2.3.2. Training procedure

Children were trained in groups of up to 10, in a specially adapted room inside the school. Every child played with a personal computer. There were 5 experimenters in the room, each evaluating 2 children. Before starting the first training session each child was individually told the instructions for their respective games.

For the PR-training task, children were told: “During the game you will see a column with a red part representing strawberry juice, and a blue part representing water. Your task will be to indicate how much strawberry flavor will the mixture of juice and water have, by clicking on this line at the bottom”. Then they were explained the significance of the end points of the ranking bar: “For example, if the mixture has a lot of juice and a little amount of water, you would say it tastes only a little of strawberry, and would respond towards the left side. If the mixture has a lot of juice and a little amount of water, you would say it tastes a lot of strawberry, and would respond towards the right side”. Then they were told that they could use the ranking bar continuously: “Remember, you can click all along the line to indicate the taste of the mixture”. Afterwards they completed 3 practice trials (proportions: 0.4, 0.6, 0.55) and were asked if they had understood the game. Only a few children required hearing the instructions a second time.

For the MC-training task, children were told: “During the game you will see a column with a red part representing strawberry juice, and a blue part representing water. Your task will be to indicate if there is more juice or water in the mixture by clicking on the right side or on the left side of this line at the bottom”. Then, they were explained how to select the correct side. “For example, if the mixture has more juice than water you would click on the right side. If it has more water than juice you would click on the left side”. They completed 3 practice trials (proportions 0.6, 0.45, 0.58). All children understood the game.

There were 5 training sessions, one per day, applied as consecutively as possible (days between session 1 and session 5: Mean = 5.98 days, Range = 5 days - 8 days). To keep children motivated with the training, we changed the type of fruit juice for each session: strawberry, pineapple, grape, orange, and strawberry again on the last day. Before starting each session, children were introduced to the new fruit juice, and were reminded of the objectives of the game. Each session lasted around 15 min. All children were tested during the same day when possible. The experimenters (who were lab volunteers) were blind to the specific hypotheses of the study. This intervention took place during regular school time but did not interfere with the normal unfolding of classes. Children were evaluated during class time (not break time) and were picked up from their classrooms in groups of 10. Teachers willingly allowed us to do this and they were blind to the specific purposes and hypotheses of our study.

### 2.4. Cognitive assessments

Before the first day and after the last day of training children completed a battery of cognitive assessments applied by trained professionals. First we describe each test and then the general evaluation procedure.

#### 2.4.1. PR near transfer - nonsymbolic proportions task

We implemented the task designed by Möhring et al. (2015) using paper and pencil, which is very similar to the PR computer training game. Stimuli appeared in a booklet with 16 different images, each
presented twice (32 stimuli in total). Each image depicted a pair of red and blue columns in stacked position at the center of the paper, representing a mixture of cherry concentrate and water. Below the columns appeared a 12 cm horizontal bar with a cartoon of a single cherry at the left end, indicating a weak taste of cherry, and a heap of cherries at the right end, indicating a strong taste. Stimuli were formed by combining units of juice with units of water, each unit being 0.5 cm height. Three practice trials were given, with proportions of juice to total amount of 0.93, 0.07, and 0.73. Test trials were formed by combining units of juice (3, 4, 5, 6) and units of total amount (juice plus water) (6, 12, 18, 24), thus giving 16 combinations (each presented twice). On each trial the child had to mark the amount of cherry taste the mixture would have along the response bar. Children did not receive feedback during test trials. Four distinct orders of the test were created for pre- and post-training evaluations, while the same items were used in each to ensure equal difficulties. For details on specific instructions see Möhring et al. (2015).

2.4.2. PR transfer - symbolic proportions

We implemented the same symbolic proportions test as Möhring et al. (2015). This is a written paper and pencil untimed test, designed for children from 3rd grade to 5th grade. It covers multiple aspects of fraction knowledge: spatial representations of fractions, ordering and equivalence relations, transcoding, and arithmetic operations. It has a total of 25 items, each is a numeric problem (i.e., without words). In order to have distinct pre- and post-training versions, we created a second, parallel form, from the original test to be used in post-training assessments. Half of the children received one form in pre-test and the other form in post-test (and vice versa). Correct responses were coded with 1 and incorrect or omitted responses with 0. Two items were not considered in the analyses for transcription errors when making the assessment – but the content of these questions was tested on other items included in the assessment.

2.4.3. Additional measures - symbolic math tests

We implemented four subtests from the Key-Math 3 Diagnostic Assessment (Connolly, 2007): Numeration, Geometry, Addition and Subtraction, and Multiplication and Division. The last two subtests are referred to as “Arithmetic” test. All these tests are paper and pencil untimed tests.

2.4.3.1. No predicted transfer – numeration

Using a booklet with pages of colorful images, presented sequentially to the subject, it measures the understanding of whole and rational numbers, by identifying, representing, comparing, and rounding numbers. It is composed of 49 items.

2.4.3.2. Predicted far transfer – geometry

Using a booklet with pages of colorful images, presented sequentially to the subject, it measures the ability to analyze, describe, classify and compare two- and three-dimensional shapes. It also covers topics such as spatial relationships and transformation, coordinates, symmetry, and geometric modeling. It is composed of 36 items.

2.4.3.3. No predicted transfer – arithmetic

The subtests “Addition and Subtraction”, and “Multiplication and Division” measure respectively the ability to perform additions and subtractions, and multiplications and divisions with whole and rational numbers. They have 35 and 31 items respectively.

The items from these four subtests are arranged in order of increasing difficulty. Correct responses were coded as 1 and incorrect or skipped responses as 0. In all tests, the evaluation stopped when a subject reached a ceiling level, defined as 4 consecutive incorrect/skipped responses. A score for each participant on each subtest was obtained by subtracting the number of incorrect/skipped responses from the ceiling level. Two parallel forms, A and B, were used for each of these subtests, with half of the children receiving form A in pre-test and form B in post-test (and vice versa).

2.4.4. No predicted transfer - approximate number system (ANS) task (computer task)

We assessed children in a nonsymbolic number comparison task, using the computer software Panamath (Halberda, Mazzocco, & Feigenson, 2008). The stimuli for this task are a set of computer images depicting a collection of blue and yellow dots. The ratio between the number of blue and yellow dots belonged to one of four bins: 1.21; 1.35; 1.56; 2.56. Children completed a total of 10 practice trials followed by around 110 test trials. To adjust for non-numerical parameters, half of the trials had equal summed area (extensive property) and half had equal individual area (intensive property). If a theorist believes that performance on ANS tasks relies on geometric/spatial representations, then they might predict transfer from PR- or MC-training to our ANS task. We believe such spatial representations play a minor role in ANS tasks and we therefore predict no transfer.

At the beginning of the task, the experimenter told the child that he/she was going to see a group of yellow and blue candies, presented side by side, and that their task was to decide if there were more yellow candies or blue candies. Children were told that the candies would appear for a short time, so they had to respond without trying to count them. For each child, the computer recorded the child's response and their reaction times. After each response children received audio feedback, consisting of a pleasant beep if the response was correct and an unpleasant beep if it was incorrect. Order of trials was randomized from pre- to post-training though the same ratios were used ensuring equal difficulty.

2.4.5. No predicted transfer - verbal abilities

2.4.5.1. Vocabulary

We applied the TEVI-test (Echeverría, Herrera, & Segure, 2002), which measures picture-to-word associative vocabulary. On each trial, children are presented with four images on one piece of paper. They listened to a word said aloud by the experimenter, and had to point to the image corresponding to this word. The test contains a maximum of 114 trials, ordered by difficulty. Correct responses were coded with 1 and incorrect with 0. Ceiling level was achieved after obtaining 6 incorrect responses among 8 consecutive items. Raw scores were computed for each participant by subtracting the number of incorrect responses from the ceiling level. Two parallel forms A and B were used, mixing half of each form in pre- and post-training assessments.

2.4.5.2. No predicted transfer - verbal analogical reasoning

As a simple measure of verbal intelligence we used the Analogy subtest from the WISC-III (Wechsler, 1997). In this test children listen aloud to a pair of words (e.g., concepts), and they have to say in what aspects they are similar to each other. For example, if the child heard “in what aspects are blue and red similar?” they might successfully reply “They are both colors”. The test is composed of 18 items ordered by difficulty. As is recommended, each response was scored as 2, 1 or 0 depending on how close it is to a set of possible responses contained in the test which have been classified according to these scores. In general, a score of 2 is given to the responses that capture the conceptual relation between the pair of words; a score of 1 is given to those responses that refer to more general relations or that are incomplete; and a score of 0 to those that do not refer to the conceptual relation. The ceiling level is determined as the first sequence of items containing 5 consecutive incorrect responses. Scores were computed for each participant by subtracting the number of incorrect responses from the ceiling level. We applied the same form in pre-and post-test evaluations.

2.4.6. Procedure for cognitive assessments

Children completed the above cognitive assessments within 2 weeks before and 2 weeks after the training. They were individually evaluated in a room inside the school. The cognitive tasks were administered in
two blocks, given on two separate days. The tasks were administered in a fixed order within a block. One block included: Numeration, Geometry, Arithmetic, and one verbal task (Vocabulary or Analogy). The other block included: Nonsymbolic Proportions, Symbolic Proportions, and the remaining verbal task.

Evaluation of ANS abilities (pre- & post-training) was conducted in another room with computers, with a preference for testing on a separate day after completing all the other cognitive tasks – but depending somewhat on each child’s class schedule (3 children completed the ANS pre-training assessment on the same day as the second block of the cognitive assessments; 52 children completed the ANS pre-training assessment on a separate day between cognitive tasks and training; and 1 child completed the ANS pre-training assessment the same day as the first training session).

3. Results

3.1. Training outcomes

We first assessed whether children improved throughout the training. For the PR-training children, we estimated the error of responses by computing the typically used measure of Percentage of Absolute Error (PAE, e.g., Fazio et al., 2014):

\[ PAE = \frac{\text{presented proportion} - \text{estimated proportion}}{\text{length of the ranking bar}} \times 100 \]

3.1.1. Proportional reasoning, PR-training children

Fig. 2 summarizes the training results for the PR-training children. To evaluate training progress, we regressed mean PAE (i.e. absolute error) on training session for each subject. Using a t-test on the slope compared to the no change value of zero we found a significant negative association between mean PAE and training session, as predicted (compared to the no change value of zero we found a significant negative error) on training session for each subject. Using a t-test on the slope of mean RT with training session, though a trend in the predicted direction was evident (t(27) = -2.213, p = .021), indicating that children produced responses with less error as the training progressed (Fig. 2A).

Following the same approach, we estimated progress in reaction time (Fig. 2B), and failed to find the predicted downward trend of the slope of mean RT with training session, though a trend in the predicted direction was evident (t(27) = -1.141, p = .264). When comparing mean RT at session 1 with the rest of the sessions collapsed we found a significant difference (note, however, that this comparison was unplanned) (t(27) = 2.213, p = .036).

Together, these results indicate that children significantly reduced their error during PR-training and, while they showed a non-significant trend towards reducing their response time across the entire training, in session 1 they were slower than in the rest of the sessions. Each of these results is consistent with improvement in the task.

3.1.2. Magnitude comparison, MC-training children

Fig. 3A summarizes the accuracy results from the Magnitude Comparison (MC-training) children, separated by ratio. The best performance in the Magnitude Comparison task results in a steep and high curve in Fig. 3A. As can be seen visually in Fig. 3A, children’s performance improved across sessions.

To estimate training progress in accuracy, we fit the psychophysical model described above to each subject and training session, and regressed the estimators of w and g over training session. In this model, a small w results in a steep slope (i.e., less internal noise allows for a better decision), and a small g results in a high curve (i.e., less guessing allows for a high asymptote). As in other studies (e.g., Odic, 2017), for one participant’s data, the model did not converge because of an extremely large g estimate (g = 1, indicating pure guessing), and another participant’s data presented an unreasonably high w (> 3). Both cases corresponded to the first training session and were not considered in the analyses. As predicted, we found a significant reduction of w over training (t(26) = -2.11, p = .045) and a significant reduction of g (t(26) = -2.718, p = .012) (see Fig. S1A in Supplemental Material).

This means that children relied on guessing less as the training progressed and they became more precise. Notice that this improvement occurred even though, in accord with our training design, children were seeing harder ratios as training progressed. This suggests that children were resilient to the increasing difficulty.

Fig. 3B summarizes the reaction time results from the Magnitude Comparison task, separated by presented ratio. Best performance would occur when children show a fast, overall RT (i.e., quick decisions), and
a steeply negative slope of RT with ratio (i.e., RTs that are sensitive to trial difficulty). As predicted, in Fig. 3B we see that both of these signatures are somewhat present, with the most pronounced improvement occurring between the first session and the rest of the sessions. To quantitatively evaluate this training improvement, we compared the rate of decrease in reaction time with ratio across the different sessions. We regressed mean reaction time over ratio for each subject and training session (see Fig. S1B), and found a significant reduction (i.e., increased negativity) in the obtained slopes with training sessions ($t$ (26) = −5.456, $p < .001$).

Overall these results show that children in the MC-training condition also improved in their performance with training: they became more accurate and faster, and made less guesses in their responses.

### 3.2. Transfer effects

In our third series of analyses, we assessed whether any training benefit transferred to the other tasks. To estimate transfer to the cognitive tasks we computed standardized gain scores for each child and task by subtracting pre- from post-training scores, and dividing by the standard deviation of all subjects in pre-training scores (Park & Brannon, 2013, 2014). Then we used $t$-tests and mixed-ANOVAs to evaluate the significance of the gains and potential differences between groups. Note that comparing gains scores between groups is statistically equivalent to testing the interaction between the factors: Time of evaluation (Pre-, Post-test) and Group in an ANOVA. Note also that we found no differences between groups on any measures at pre-training (all $p > .11$). We also included Bayesian analyses to assess the evidence for the null hypothesis of no differences between groups when it was of theoretical importance (Raftery, 1995; see Masson (2011) for a practical tutorial on this). Each statistical test was evaluated at the $p < .05$ level of significance. We corrected the significance level of alpha (0.05) using Bonferroni criteria when testing either a small set of exploratory post-hoc hypotheses or a full range of post-hoc pairwise comparisons. Effect sizes for the ANOVAs were estimated using generalized eta squared ($\eta_p^2$), as recommended for repeated measures designs (see Bakeman, 2005). Greenhouse-Geisser corrected $P$ values were presented when sphericity assumptions did not hold.

For each cognitive task, we assessed transfer effects both by testing whether each training group significantly improved after the training (i.e. gain scores above zero) and by testing if there were differences in the observed gains between groups. We reported standardized gain scores throughout the text, having shown that training groups did not differ in Pre-training scores. The use of standardized units facilitates the estimation of size effects and is consistent with other studies (e.g. Park & Brannon, 2014).

#### 3.2.1. Proportional reasoning (paper-and-pencil) tasks

**3.2.1.1. Nonsymbolic proportions.** This task assessed the possibility of near transfer. We computed standardized gain scores in Percentage of Absolute Error (PAE) for each subject, consistent with previous studies (e.g. Fazio et al., 2014). We expected the PR-training children to display negative change in PAE (i.e., reducing their error) (because their training focused on these abilities), and the MC-training children to not show negative change in this task (because their closely related training did not focus on proportions). Consistent with these predictions (Table 1), we found a significant improvement for the PR-training children (Fig. 4) ($M = −0.75, SD = 1.065$) ($t$ (25) = −3.589, $p = .001$), whereas for the MC-training children we actually observed a significant reduction in performance (increase in error) after the training ($M = 0.384, SD = 0.972$) ($t$ (26) = 2.053, $p = .05$). And, the gain/losses (i.e. change scores) were significantly different for the two groups ($t$ (51) = −4.05, $p < .001$) with children from the PR-training improving more after training than the MC-training children.

These results suggest that the PR-training was effective in improving nonsymbolic proportional reasoning in the PR-training group. The MC-training children, on the other hand, produced an unexpected worsening effect. Although further analyses did not indicate a clear source for this effect (see Discussion), it is plausible that experience in the MC-training task may have adversely affected some procedural aspect of the nonsymbolic proportion task.

**3.2.1.2. Symbolic proportions.** This task assessed the possibility of transfer from nonsymbolic to symbolic proportions. To estimate transfer effects, we computed standardized gain scores from the percentage of correct responses. We predicted positive gain scores for the PR-training group and non-significant gain scores for the MC-training group (Table 1). Fig. 4 depicts gain scores on this test across all question types. Planned comparisons to the chance level (zero gain) revealed significantly positive gain scores for the PR-training group ($M = 0.295, SD = 0.692, t$ (25) = 2.172, $p = .040$), and no significant gains for the MC-training group ($M = 0.057, SD = 0.83, t$ (26) = 0.355, $p = .725$). However, these differences must be taken with caution as the more stringent comparison of the gain scores for the two groups did not reach significance ($t$ (51) = 1.131, $p = .263$). We explore the potential lack of power for this comparison in our replication experiment below.

Taken together, these results suggest some transfer effects to symbolic proportional reasoning.

**3.2.2. Geometry and arithmetic tests**

Another important goal in our study was to investigate causal connections from our training tasks to other math-related abilities. We chose...
arithmetic and geometry, as recent studies suggest these abilities may be built from different nonsymbolic foundations (e.g., Lourenco et al., 2012). Specifically, it is thought that reasoning with continuous quantities contributes to the development of geometry knowledge, while reasoning with discrete quantities contributes to the development of arithmetic knowledge (Dehaene et al., 2006; Hyde et al., 2014; Lourenco et al., 2012).

Recall that in our training program we used continuous quantities in both tasks. So, if reasoning with continuous quantities is specifically and causally related to geometry, we would expect improvements in geometry and no improvements in arithmetic. Indeed, a 2 Test (Geometry, Arithmetic) by 2 Group (PR-training, MC-training) mixed ANOVA showed a significant effect of Test, as Geometry gain scores (M = 0.327, SD = 0.803) were higher than Arithmetic scores (M = −0.073, SD = 0.531) (F(1, 53) = 10.02, p < .05). While both training PR- and MC-groups showed positive gains in Geometry (Fig. 5), these must be qualified somewhat as they reached significance for the MC-training children (M = 0.377, SD = 0.849, t (26) = 2.308, p = .029), and only close to significance for the PR-training children (M = 0.279, SD = 0.769, t(27) = 1.918, p = .066). Using a Bayesian procedure (Masson, 2011), we found that the probabilities of the null hypothesis (no difference between groups) and the alternative hypothesis were respectively Pnull(H0|D) = 0.87 and Pnull(H1|D) = 0.13, corresponding to positive evidence of no difference between groups (Raftery, 1995). Together, these results suggest, as predicted, that training with continuous magnitudes shows significant far transfer to geometry.

Regarding Arithmetic, considering both subtests together (Addition &Subtraction and Multiplication&Division) we found the predicted nonsignificant gain scores for the MC-training children (M = 0.072, SD = 0.575, t(26) = 0.644, p = .525), and significant worsening for the PR-training children (M = −0.212, SD = 0.451, t(27) = −2.483, p = .02). There was also a significant difference between the groups (t (53) = −2.035, p = .047). For the Addition and Subtraction test we found no gains in the MC-training group (M = 0.018, SD = 0.637, t (26) = 0.143, p = .887) (Fig. 5), with a trend in the opposite direction in the PR-training group (M = −0.203, SD = 0.568, t(27) = −1.89, p = .069); also neither group showed gains in the Multiplication and Division subtest (MC-training: M = 0.137, SD = 0.641, t(26) = 1.109, p = .278; PR-training: M = −0.172, SD = 0.546, t(27) = −1.662, p = .108). Thus, intuitive training on continuous magnitudes does not transfer to symbolic arithmetic performance, and PR-training may even worsen performance on this test – though exploring the possible causes of these changes was not part of our design.

3.2.3. Numeration and approximate number system tasks

As predicted, neither group showed gains in the Numeration test – our test of children’s general conceptual understanding of numbers (PR-training children: M = 0.136, SD = 0.544, t(27) = 1.324, p = .197; MC-training children: M = 0.141, SD = 0.643, t(26) = 1.142, p = .264) (data not shown).

Regarding the nonsymbolic numeric representations task (ANS), as predicted for both training groups, we did not observe significant gains in accuracy (PR-training children: M = −0.157, SD = 1.573, t (25) = −0.511, p = .614; MC-training children: M = 0.193, SD = 1.024, t(25) = 0.957, p = .348). However, we found significant improvements in reaction time for both groups (PR-training children: M = −0.822, SD = 0.753, t(25) = −5.569, p < .001; MC-training children: M = −0.693, SD = 0.748, t(25) = −4.725, p < .001), showing that children did get faster at this task. There were no differences between groups in accuracy and reaction time gains (accuracy: t (50) = −0.951, p = .36; reaction time: t(50) = 0.62, p = .538). These gains in reaction time mean that children were able to make their decisions significantly faster in the post-training ANS test than in the pre-training test, while still maintaining their same level of accuracy. These results must be qualified as these RT improvements might suggest either a positive impact of continuous magnitude training on discrete magnitude comparison and/or test-retest improvement.

3.2.4. Verbal tests

As predicted, neither group showed gains in Vocabulary or Verbal Analogies – tests, which were included to assess if our interventions would simply improve cognition and performance more generally (PR-training children: Vocabulary: M = −0.321, SD = 1.334, t(27) = −1.272, p = .214; Analogies: M = 0.217, SD = 0.728, t(25) = 1.519, p = .141; MC-training children: Vocabulary: M = 0.107, SD = 1.143, t (26) = 0.487, p = .63; Analogies: M = 0.136, SD = 0.549, t(26) = 1.284, p = .211). These results support the specificity of our intervention to math abilities.

4. Experiment 2

The results of our first experiment were broadly consistent with our predictions that PR-training could improve PR performance, show near
transfer to non-symbolic proportions, transfer to symbolic proportions and far transfer to geometry. Our sample size per group (≈27) was sufficient to demonstrate these effects, but one might rightly worry that this experiment was under powered, and that our results may be spurious. To assess the replicability of our previous results, specifically concerning proportional reasoning trainability and transfer effects, we recruited an independent group of 4th grade children and trained them in a similar nonsymbolic proportional reasoning task. We used the same pre-post-training design and cognitive assessments, with any minor differences indicated below.

4.1. Material and methods

4.1.1. Participants

Thirty-three fourth-grade children (16 Female, Mean Age: 9.3 years, range: 9–10 years) participated in the study, and were recruited from a public school in Santiago, Chile. Three participants completed only one or two training sessions, and so were not considered in the analyses. Among the rest of the children, 27 completed both pre-and post-training evaluations of all cognitive tests, with the exception of the Flanker Task, which was completed by 22 children. Children and their parents/caregivers gave written consent before participating in the study. At the end of the study children received a small gift for their participation. The study was approved by the local ethics committee.

4.1.2. Tasks and procedures

4.1.2.1. PR-Training game. We used a new game in Experiment 2. As in Experiment 1, the goal of the game was to exercise children’s ability to estimate relative areas. In each trial, children were exposed to two sets of colored dots of the same size (homogenous arrays), one on top of the other. Their task was to estimate the dot size of one array relative to the dot size of the other array, and to translate that estimation to a horizontal ranking line (following the same logic as the PR-training task in Experiment 1) (Fig. 6). The array below was the reference. For example, if the array below presented dots with 20 pixels² of individual area, and the array above presented dots with 80 pixels² of individual

Fig. 5. Transfer to Geometry and Arithmetic abilities. ‘Add&Sub’ and ‘Mult&Div’, ‘Addition and Subtraction’ and ‘Multiplication and Division’ subtests from Written Arithmetic (*, p < .05).

Fig. 6. Training task and training outcomes for Experiment 2. A, sample image from the training task. Children estimated the area of yellow dots relative to the area of blue dots. In this example a proportion of 0.3 (see the green region of the feedback rectangle). To help children remember the task, two cartoons on either side of the response line depicted anchor values: a small yellow and a large blue dot on the left side (to indicate a small proportion value), and a big yellow and a small blue dot were on the right side (to indicate a larger proportion value; see main text for further details). B, mean PAE across sessions, showing that children became more accurate at the task as training progressed. The line is the best linear fit at the group level. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
area, this would correspond to a relative size of 0.2. Children were instructed to ignore the number of dots and to focus only on their sizes. Trial onset was controlled by the participant, who initiated each trial by pressing the space bar. The dots remained onscreen until the participant responded, with a timeout of 5000 ms (see Supplementary Material for a video of the task). Note that although the stimuli were formed by discrete elements (i.e. dots) the quantification involved a continuous dimension (i.e. size), thus making it comparable to the PR-training task. We included this change from Experiment 1 training in order to extend our results to discrete displays.

Each stimulus was created by choosing a target proportion from among 9 levels (0.05, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 0.95, jittered by a small amount to add variability). Then the number of dots for each array was randomly chosen from the range of 1–50. As mentioned, the dots of the same color had the same individual sizes and were randomly distributed inside invisible rectangles (avoiding overlap). Stimuli were generated using a Matlab script developed by Dehaene, Izard, and Piazza (2005), adapted for our specific purposes.

In each training session, we presented the 9 proportion levels, 4 times, yielding 36 trials, divided in two blocks of 18 trials. Children received audio-visual feedback and were also shown a running summary of their performance, exactly as in the Experiment 1 PR-training program. Across sessions, the color of the dot arrays varied to keep children engaged, combining yellow, blue, red, green and orange.

4.1.2.2. Training procedure. At the start of the training, children were shown a picture with yellow and blue dots and were told: “This is a party, with yellow and blue friends. Sometimes, the yellow friends are bigger than the blue ones, and sometimes they are smaller. Your task will be to indicate the size of the yellow friends in relation to the size of the blue friends. Remember, we will not care about how many friends there are, only about their sizes”. Then, they were shown sequentially two pictures that served as anchor points (depicting proportions of 0.95 (i.e., yellow bigger than blue) and 0.05 (i.e., yellow smaller than blue)). The experimenter said: “For example, in this party, the yellow friends are very big, and the blue friends are small, so we mark on this side” (the evaluator pointed to the right side of the ranking line where a vertical line indicated the correct position). Then: “In this other party, the yellow friends are small, and the blue ones are very big. Where would you put the mark in this case?”. After children pointed to some position (typically to the left side) they received feedback (a vertical line in the correct position on the ranking line). Finally, to show children that they could use the ranking bar continuously, they were presented with a picture showing dots of equal size between the arrays (a proportion of 0.5). Almost all children spontaneously pointed to the half of the line and received feedback.

After the instructions, children received a computer practice session of 10 trials with feedback. Then the first training session started. Children were reminded of the instructions briefly before starting subsequent sessions (without practice). We recorded the response position and reaction time.

4.1.3. Cognitive assessments

Children completed the same battery of tests as in the Experiment 1, with the following exceptions: we did not include the Numeration and the ANS tasks; for the control tasks we excluded Vocabulary but included a measure of inhibitory control (flanker task). For this inhibition task we followed the computer procedure by Matthews et al. (2016). Briefly, each trial began with the presentation of a fixation cross (500 ms) followed by the presentation of a stimulus composed of 5 arrows (e.g. "< < < > < < "). The task was to indicate the direction of the central arrow. Stimuli remained on the screen until the participant responded with a timeout of 2000 ms. There were 15 practice trials followed by 80 test trials. We used the proportion of correct responses as a performance measure on this task.

4.2. Results

4.2.1. Training outcomes

Following the same approach to quantify training progress as in Experiment 1, we found that children improved at the task, producing responses with less error (mean PAE) across sessions (Fig. 6) (t(29) = −7.069, p < .001). Reaction time did not show variations over sessions (data not shown) (t(29) = −0.435, p = .667), consistent with no trade-off between accuracy and speed of response. These results are in agreement with what we observed in Experiment 1, and support our prediction that intuitive proportional reasoning is trainable in children.

4.2.2. Transfer effects

4.2.2.1. Proportional reasoning (paper-and-pencil) tasks. As in Experiment 1, we found significant improvements in both nonsymbolic and symbolic proportions tasks (Fig. 4) (Nonsymbolic: M = −0.552, SD = 0.833; t(26) = −3.444, p = .002 Symbolic: M = 0.602, SD = 1.342; t(26) = 2.329, p = .03), replicating our previous findings.

4.2.2.2. Geometry and arithmetic tests. Replicating the results of Experiment 1, children showed far transfer to Geometry and no transfer to the Arithmetic test (Fig. 5). In Experiment 2, the improvements in Geometry were significant (M = 0.337, SD = 0.821; t(26) = 2.132, p = .043). For the Arithmetic test, as in Experiment 1, we found a significant worsening (M = −0.319, SD = 0.785; t(26) = −2.114, p = .044), though we do not know why this result should occur. For each subtest separately we found the same trend as in the PR-training (Fig. 5) (Addition & Subtraction: M = −0.302, SD = 1.007; t(26) = −1.558, p = .131; Multiplication & Division: M = −0.23, SD = 0.839; t(26) = −1.427, p = .166).

4.2.2.3. Control tasks (verbal analogies and flanker task). Consistent with the results of Experiment 1, we found no gains in Verbal Analogies (M = 0.262, SD = 0.562; t(26) = 0.26, p = .797). However, performance in the flanker task showed improvements after training (M = 0.364, SD = 0.434; t(21) = 3.938, p < .001) (data not shown). This suggests that our training task may also show transfer to executive function skills. Because such abilities are quite general and relied on in many tasks, this is a possible mechanism for transfer in our other tasks – which would be consistent with some theories (Cragg & Gilmore, 2014; Gómez, Jiménez, Bobadilla, Reyes, & Dartnell, 2015).

5. Discussion

To summarize the main results from our study, we show that 1 week of training on a simple computerized nonsymbolic proportional reasoning task shows near transfer to a paper-and-pencil nonsymbolic proportional reasoning task, transfer to symbolic proportional reasoning, and far transfer to geometry. These results were consistent across two different experiments. An active control group trained in comparing continuous extents did not show improvements in the proportional reasoning tasks and, as predicted, did show far transfer to Geometry. Also as predicted, neither group improved in symbolic arithmetic, nor in two verbal tasks, although some gains were also observed in inhibitory control in Experiment 2. Overall these results suggest specific as well as shared contributions of nonsymbolic proportional reasoning and magnitude comparison to symbolic math: while the understanding of symbolic fractions appears more closely linked to intuitive proportion computations, both quantitative processes would contribute to geometry knowledge.

Our work complements a growing body of studies addressing the role of nonsymbolic proportional reasoning in symbolic math (Fazio et al., 2014; Matthews et al., 2016; Möhring et al., 2015). These studies have reported positive correlations between nonsymbolic proportional abilities and knowledge of symbolic fractions. Here we extended these
observations in two significant ways. First, by showing that the intuitive understanding of proportions is trainable in children, which suggests that the malleability of nonverbal quantification systems is not restricted to absolute magnitudes, but also includes the computation of relational quantities. And second, by investigating the impact of this training experience on a wider range of math abilities, providing insight into specificity of training quantitative intuitions and their transfer to other aspects of math knowledge.

A potential concern of our study is that children in each training group misunderstood their respective tasks (PR-training children performed magnitude comparison and/or MC-training performed proportional reasoning), due to the high perceptual similarity between them. However, training results as well as complementary analyses suggest that children did understand their tasks.

Specifically, reaction time analyses showed that for the MC-training group reaction time increases when proportions approach to 0.5 (or equivalently, the ratios approach to 1, see Fig. 3), which is consistent with a discrimination/comparison process that becomes harder as stimuli get closer, a behavior widely observed in the discrimination literature (e.g., Dehaene, 2007). Importantly, PR-training children displayed the opposite pattern, that is, reaction time tend to decrease as proportion values approach to 0.5 (see Supplemental Fig. S2). The latter is consistent with a relative easiness of judging the half compared to other proportions, as reported in proportional reasoning studies (e.g., Hollands & Dyre, 2000; Spence & Krizel, 1994). On the other hand, we also observed that the scaling factor (the ratio between the height of the central columns and the length of the ranking line) influenced accuracy in the PR-training group (the higher the scaling factor the higher the error, replicating Möhring et al., 2015, data not shown). However, this was completely absent in the case of MC-training children (Fig. S3); instead accuracy was modulated by ratio values, consistent with Weber's law for discrimination of magnitudes (Fig. 3). These results thus suggest that children in both groups correctly perceived their respective games.

Another caveat of our study is that the group of children trained in magnitude comparison (MC-training) decreased performance in the paper-and-pencil nonsymbolic proportional reasoning task after training. One possibility to explain this effect was that the framing of the MC-training task (in which we placed a horizontal bar bellow the columns stimuli to increase perceptual similarity with the PR-training task) induced some bias for responding towards the extremes in the nonsymbolic proportion task (i.e. more to the left for proportions < 0.5 and more to the right for proportions higher than 0.5). However, when evaluating the changes in the amount of bias before and after the training in the MC-training children, we observed no significant differences (see Fig. S4 and the statistics therein). Moreover, we also found no association between improvements in the MC-training task for MC-training children (quantified as changes in the psychophysical parameters w and g, see above) and gains in the nonsymbolic proportion task (for w and gains in PAE: r = 0.121, p = .549; for g and gains in PAE: r = −0.071, p = .724). This suggests no trade-off between magnitude comparison and proportional reasoning. Furthermore, it also seems unlikely that the MC-training dampened general motivation or attention since the worsening effect was rather specific, affecting only the nonsymbolic paper-and-pencil proportions test. Nevertheless, although the framing of the MC-training had no effect on the bias of responses, as just mentioned, it is plausible that it may have affected some procedural aspect of the nonsymbolic proportion task. Other training studies have also reported worsening effects without an evident cause (see Khanum et al. (2016), who reported that training in adding line lengths had a worsening effect on the number line task). Future studies may contrast the effects of framings, and more generally, determine the training conditions that tackle specific cognitive abilities while keeping others unchanged.

Regarding our main question concerning the specific causal link between nonsymbolic and symbolic proportional reasoning, in two experiments, we obtained consistent evidence supporting it. However, the lack of significant differences between PR- and MC-training groups in Experiment 1 for the symbolic fraction test may suggest that the test we used was not sensitive enough to expose this difference. Another possibility is that the link between nonsymbolic and symbolic proportional reasoning is not so strong and specific in primary-school children, when they are beginning to master the symbolic representations of fractions; at least not so strong that a purely nonsymbolic proportional reasoning training would be enough to enhance performance in symbolic fractions compared to a training in magnitude comparison at this age, but possibly later in time. Consistent with this possibility, a recent study of nonsymbolic arithmetic training in pre-school children (Dillon et al., 2017), showed no transfer effects to symbolic math compared to an active control group, but similar training reports have shown positive transfer effects in older kids (1st grade) (Hyde et al., 2014; Khanum et al., 2016; Wang et al., 2016) and adults (Park & Brannon, 2013; Park & Brannon, 2014). It has been proposed that experience in manipulating symbolic representations of numbers may help strengthen the association with nonsymbolic conceptual referents and generate specificity in the connections between different types of quantitative representations (Bonny & Lourenco, 2015; see also Piazza, Pica, Izard, Spelke, & Dehaene, 2013). As these associations become more robust and specific, the power to detect specific transfer effects may increase too. Whether this is the case for proportional reasoning is something that future studies will elucidate.

On the other hand, our study also contributes to an understanding of how nonsymbolic proportional reasoning may interface with other branches of school-learned mathematics besides symbolic fractions, namely to geometry and symbolic arithmetic. The similar pattern of gains observed in both training groups, in both experiments – i.e., geometry was improved but symbolic arithmetic was unchanged or even reduced – suggests a more functional link between continuous extent training (PR- and MC-training) and geometry. This result must be qualified, however, by the successful transfer to executive function abilities in Experiment 2 – that is, the training of executive functions might be a plausible mechanism for the general improvements in geometry that we saw across all groups.

Regarding symbolic arithmetic, the null or negative gains observed for both training groups are in line with studies reporting that nonsymbolic discrete quantities, but not continuous ones, are linked with symbolic arithmetic. For instance Hyde et al. (2014) showed that training in performing comparison or additions with collections of dots – but not additions of lines or brightness comparisons – transfers to symbolic arithmetic in 1st grade children. Similarly, Lourenco et al. (2012) reported specific associations between nonsymbolic number acuity and arithmetic, but not between nonsymbolic area acuity and arithmetic in college students. Indeed, neither PR- nor MC-training children displayed gains in ANS acuity only in RT), nor in the Numeration test, suggesting multiple failures to observe transfer from continuous extent training to arithmetic performance in our children.

Taken together, the pattern of results across two experiments is consistent with our predictions and suggests some specific transfer to nonsymbolic and symbolic proportional reasoning for the PR-training group and a broader far transfer to geometry for both the PR- and MC-training groups.

Overall, and as a more general implication of our data, the results in the geometry and arithmetic tests suggest that different branches of symbolic mathematics may interface differentially with continuous and discrete nonsymbolic magnitudes. Thus, the proposal that magnitudes are processed through a generalized magnitude system (Walsh, 2003; for reviews see Cohen Kadosh, Lammertyn, & Izard, 2008; Henik, 2016), at least in its strong formulation, would not be consistent with our data.

An important question for future research is to determine the nature of the link between nonsymbolic and symbolic proportional reasoning. One possibility is that both rely on an abstract representation of relative
magnitude. Neural data gives support to this idea. Using an fMRI adaptation paradigm, Lewis, Matthews, and Hubbard (2015) showed that repeated presentations of line pairs depicting some ratio (e.g. 3/10), led to a subsequent rebound of the BOLD signal in a distance-dependent manner by the presentation of a deviant ratio. Crucially, these effects were seen when the deviant was either a symbolic fraction or a ratio of lines. The activated brain region was the right mid-IPS (see also Vogel, Grabner, Schneider, Siegler, & Ansari, 2013).

Another mechanism is that nonsymbolic and symbolic proportional reasoning both focus on relational quantities. Indeed, one of the typical errors children make when dealing with proportions is to focus only on numerators or denominators when comparing fractions, a tendency even present in contexts that do not use numerical symbols but where children can count and match discrete stimuli (Boyer, Levine, & Huttenlocher, 2008). It has been suggested that this may explain why children perform much better in proportional tasks that use continuous rather than discrete objects (Boyer et al., 2008). Interestingly, Boyer and Levine (2015) reported that a brief practice in performing continuous proportional reasoning led to subsequent improvements in the same proportional task with discrete stimuli in young children. Similar to our study, this suggests that engaging continuous proportions may help children to understand that proportions are relations between two components and so mitigate children’s tendency to focus on just one part of this relation.

In conclusion, our work has shown that a simple computerized training program can improve children’s intuitive proportional reasoning and provides suggestive evidence for a functional relation to symbolic fractions knowledge in children. We hope the rational, design, and results of our study may motivate and help guiding future research.

Authors’ contribution

Camilo Gouet: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing-original draft, Writing-review & editing, Project administration, Funding resources.

Salvador Carvajal: Software, Resources, Writing-review & editing.

Justin Halberda: Conceptualization, Methodology, Software, Resources, Writing-original draft, Writing-review & editing, Funding resources.

Marcela Peña: Conceptualization, Methodology, Resources, Writing-review & editing, Supervision, Funding resources.

Acknowledgements

We thank all the parents and children who participated in the study. We also thank to Villa Macul School’s staff and Sylvia Campos for contacting us to the school. We acknowledge the Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), Chile, for funding resources through the fellowship #21130131, and grant Fondecyt #1141040. We acknowledge the Latin American School for Education, Cognitive and Neural Sciences (LASchool). We thank the useful commentaries of reviewers. Thanks to Tyler Knowlton for data analysis assistance and to Josh Langfus for helpful discussions and for writing and data analysis assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.104154.

Appendix B. Supplementary information


