



Uncovering the interplay between drawings, mental representations, and arithmetic problem-solving strategies in children and adults

Hippolyte Gros¹ · Jean-Pierre Thibaut² · Emmanuel Sander³

Accepted: 14 January 2024 / Published online: 12 February 2024
© The Psychonomic Society, Inc. 2024

Abstract

There is an ongoing debate in the scientific community regarding the nature and role of the mental representations involved in solving arithmetic word problems. In this study, we took a closer look at the interplay between mental representations, drawing production, and strategy choice. We used dual-strategy isomorphic word problems sharing the same mathematical structure, but differing in the entities they mentioned in their problem statement. Due to the non-mathematical knowledge attached to these entities, some problems were believed to lead to a specific (cardinal) encoding compatible with one solving strategy, whereas other problems were thought to foster a different (ordinal) encoding compatible with the other solving strategy. We asked 59 children and 52 adults to solve 12 of those arithmetic word problems and to make a diagram of each problem. We hypothesized that the diagrams of both groups would display prototypical features indicating either a cardinal representation or an ordinal representation, depending on the entities mentioned in the problem statement. Joint analysis of the drawing task and the problem-solving task showed that the cardinal and ordinal features of the diagrams are linked with the hypothesized semantic properties of the problems and, crucially, with the choice of one solving strategy over another. We showed that regardless of their experience, participants' strategy use depends on their problem representation, which is influenced by the non-mathematical information in the problem statement, as revealed in their diagrams. We discuss the relevance of drawing tasks for investigating mental representations and fostering mathematical development in school.

Keywords Context effects · Drawing · Mathematical cognition · Mental models · Problem solving

Introduction

Word problem solving is a central component of mathematics education (Verschaffel et al., 2020). Calculating how many eggs are needed to cook an omelet or how long it takes for a bathtub to fill up are typical exercises designed to help children learn how to use abstract mathematical notions in concrete, real-life situations. But how exactly does one proceed to translate a series of words and sentences describing a specific situation into an algorithmic procedure leading to the solution? A growing line of research suggests that

problems are encoded into a mental representation including both mathematical and non-mathematical information, which is then translated into a solving algorithm (Gros et al., 2020b). Thus, non-mathematical information about the situation described in a problem statement would influence its mental representation, and learners' ability to find an appropriate solving strategy would depend upon its semantic content (Gros et al., 2019).

This prediction has led to a new way of investigating a foundational distinction in mathematics: the difference between the ordinal property of numbers (their rank in an ordered list) and their cardinal property (the number of elements in a set). More specifically, it has been shown that the mention of specific daily-life entities within a problem statement could be enough to lead learners to construct either an ordinal mental representation of the numerical situation, or a cardinal representation instead (Gamo et al., 2010; Gros et al., 2021). For instance, mentioning elevators moving between floors would make it more likely both for lay adults and mathematicians to construct an ordinal

✉ Hippolyte Gros
hippolyte.gros@cyu.fr

¹ CY Cergy Paris University, Paragraphe Lab, Gennevilliers, France

² Université de Bourgogne, LEAD-CNRS UMR 5022, Dijon, France

³ University of Geneva, IDEA Lab, Geneva, Switzerland

encoding underlining the order between the different floors. On the other hand, mentioning a collection of marbles being counted would lead instead to a cardinal encoding of the situation, focusing for instance on color sets (Gros et al., 2021).

However, these representational differences have only been investigated through indirect measures, such as strategy use (Gros et al., 2021), response times (Gros et al., 2019), or saccadic eye movements (Gros et al., 2020a). Considering the ongoing debate regarding the nature of the mental representations elicited by arithmetic word problem solving (Bassok, 2001; Daroczy et al., 2015; Gros et al., 2020b; Gvozdic & Sander, 2020; Orrantia & Múñez, 2013; Thevenot, 2010; Thevenot & Barrouillet, 2015; Verschaffel et al., 2020), it seems crucial to gather additional evidence, probing more directly into the structure of individuals' mental representations of arithmetic word problems. In this study, we proposed to investigate the use of drawing production as a window into the mental representations of arithmetic problems and to analyze their relationship with strategy choice.

Investigating mental representations through drawing productions

The idea that different representations are abstracted depending on the semantic content of a problem statement is compelling in that it provides an account of some key performance differences reported in the literature (e.g., Coquin-Viennot & Moreau, 2003; De Corte et al., 1985; Gamo et al., 2010; Hudson, 1983; Martin & Bassok, 2005; Thevenot & Oakhill, 2005; see Gros et al., 2020b, for a discussion of this argument). However, evaluating the precise structure of these representations necessarily requires taking an indirect route, since direct investigation of mental constructs is seldom possible. Over the years, cognitive scientists have employed a variety of techniques to study individuals' mental representations (Pearson & Kosslyn, 2015). Classical examples of indirect measures include reaction times (e.g., Shepard & Metzler, 1971), verbal reports (e.g., Ericsson & Simon, 1980), self-assessment questionnaires (e.g., Weinman et al., 1996), metaphors (e.g., Lakoff & Núñez, 2000), gestures (e.g., Fuhrman & Boroditsky, 2010), written statements (e.g., Pinnegar et al., 2011), eye movements (Fourtassi et al., 2017), fMRI activation patterns (Lewis-Peacock et al., 2015), event-related potentials (Bagnoud et al., 2018), or typicality ratings (Hebart et al., 2020), to name only a few. However, it can be argued that none of these metrics have the high-dimensional complexity and richness of drawings when it comes to investigating one's mental representations (Bainbridge, 2022). In this paper, we intend to show that drawing analysis can help us gain a deeper understanding of some crucial differences in the representation of arithmetic word problems.

Drawing tasks have been regularly used in the history of psychology, notably to perform clinical diagnoses (Agrell & Dehlin, 1998; Makuuchi et al., 2003; Shulman, 2000; Wechsler, 2009; see Gainotti & Trojano, 2018, for review), or to study the affective processes of children and young adults (Burgess & Hartman, 1993; Silver, 2009). The visual complexity of diagrams has also made it possible to investigate object and scene representations through drawing tasks (e.g., Chamberlain & Wagemans, 2016; Freeman & Janikoun, 1972; Kosslyn et al., 1977; see Bainbridge et al., 2019, for a review). While some authors have raised questions regarding the difficult interpretation of complex and subjective drawings in previous studies (Thomas & Jolley, 1998), recent years have seen a resurgence of interest in drawing-based research, with an increasing number of studies being conducted and a renewed recognition of its significance among the research community (Bainbridge et al., 2019; Bainbridge, 2022; Long et al., 2018).

Overall, drawing production has been shown to be a promising path to investigate the structure of one's representation without resorting to explicit verbalization, especially among children (Bainbridge, 2022). A well-known example comes from Vosniadou and Brewer's (1992) seminal study on conceptual change, in which they elicited drawings from third and fifth grade children to study the development of their mental representation of the earth. By asking them "can you draw a picture of the earth?" and a few follow-up questions such as "now draw the sky" or "show me where the moon and stars go," they were able to differentiate between, for example, children adopting a "flattened sphere" earth model, children adopting a hollow sphere model, and children adopting a rectangular earth model. Several studies have since resorted to drawing tasks to study conceptual change among children and adults (e.g., Hobson et al., 2010; Mikkilä-Erdmann et al., 2012; Trundle et al., 2007; Ucar et al., 2011).

When it comes to mathematics, the use of drawings has been a longstanding informal practice for teachers to gain insight into children's conceptual development (Crespo & Kyriakides, 2007), yet a comparatively limited number of studies have been conducted using systematic drawing analysis (Carruthers & Worthington, 2003). Previous works have notably used diagram production tasks to look at the understanding of geometrical notions (e.g., De Bock et al., 1998, 2003; Thom & McGarvey, 2015), as well as fraction representations (e.g., Tunç-Pekkan, 2015; Westenskow et al., 2014; Yoshida & Shinmachi, 1999). In the field of mathematical problem solving, Cummins (1991) conducted one of the first studies using drawing tasks to investigate children's interpretation of arithmetic word problems. In two experiments, she probed first-grade children's interpretation of a series of additive word problems involving marbles, based on Riley et al.'s (1983) problem typology. She asked

the participants to solve the problems and then to draw a representation of the marbles in each problem. Her results suggested that drawing accuracy was significantly correlated to solving performance. Other works have found converging evidence that drawing accuracy may predict solving performance in mathematics (De Bock et al., 1998; Uesaka et al., 2010; Van Essen & Hamaker, 1990).

Since this seminal work, a number of studies have resorted to drawing tasks to study the mental representation of mathematical problems (e.g., Bakar et al., 2016; Barrios & Martínez, 2014; Csíkos et al., 2012; Edens & Potter, 2007, 2008; Reeve, 1996; Rellensmann et al., 2017). For instance, Reeve (1996) used a diagram production task to investigate children's conceptual understanding of fractions. He analyzed to what extent the drawings of Grade 7 and 8 students conformed with relevant mathematical principles. He found that drawing adherence to the corresponding mathematical rules was predictive of their problem-solving performance. In a similar perspective, Edens and Potter (2008) instructed fourth and fifth graders to solve an arithmetic word problem and to make a drawing to help them find the solution. Using a custom scale, they graded the extent to which the students' drawings were schematic or pictorial. They showed that the construction of schematic drawings was positively correlated with solving performance, and that most students (79%) rendered schematic representations. In a similar perspective, other studies have found evidence that the degree of abstraction of an individual's drawing is correlated with their problem-solving performance (Hegarty & Kozhevnikov, 1999; Rellensmann et al., 2017; Van Garderen & Montague, 2003). It should be noted that while these previous studies have been able to identify a link between general drawing qualities and solvers' performance in mathematical word problem solving, there appears to be a gap in the literature regarding the relationship between the features of a drawing, the content of a mental representation, and the choice of a solving strategy. The current study intends to tackle this question by asking participants to solve and draw problems admitting several distinct solving strategies.

Insights on problem encoding from strategy choice

Indeed, in addition to drawing analysis, another promising path to study learners' representation of arithmetic word problems comes from the study of strategy choice in problems admitting multiple solving strategies. For instance, Thevenot and Oakhill (2005, 2006) worked on a multiple-step problem-solving task in which the cognitive load was manipulated through the range of the problem's values (using either two-digit or three-digit numbers). They showed that depending on the magnitude of the values, participants used different solving algorithms, thus suggesting that a specific representational change had occurred on high-value

problems. Similarly, when studying arithmetic word problems, the selection by the participants of one solving strategy over another can yield valuable insights into the constructed mental representation (De Corte et al., 1985).

In fact, according to the SECO (Semantic Congruence) model (Gros et al., 2020b), strategy choice is directly dependent upon the structure of the mental representations constructed while attempting to solve a problem. SECO predicts that the encoding of arithmetic word problems is significantly influenced by the problem's world semantics (the non-mathematical, daily-life knowledge evoked by the entities described in the problem statement). This leads individuals to construct a mental representation in working memory, the features of which depend not only on the mathematical information in the problem, but also on the non-mathematical information attached to the problem statement. SECO predicts that this representation dictates which solving strategy can be used by the solver. In other words, the non-mathematical information in the problem statement may constrain the problem representation and lead to one strategy being used over another. Thus, by using word problems admitting several distinct solving strategies, it should be possible to explore the structure of the underlying mental representation and to pinpoint the semantic constraints influencing it (Gros et al., 2020b).

A straightforward example of this idea comes from Coquin-Viennot and Moreau's (2003) study on multiplicative problems. They created word problems admitting two solving strategies: either a factorization algorithm (e.g., " $14 \times (5 + 7)$ ") or a more costly development algorithm (e.g., " $14 \times 5 + 14 \times 7$ "). When the problem statement mentioned different sets of flowers being counted, participants tended to use the development procedure. However, when the problem statement also mentioned that the flowers were put in sets to form bouquets, a higher number of participants were able to use the factorization strategy instead. More recent studies have also used dual-strategy distributive problems to assess students' conceptual knowledge (Scheibling-Sève et al., 2020, 2022). In their study, Scheibling-Sève et al. (2020) recorded which strategy fourth and fifth graders used to solve isomorphic word problems. Their strategy-choice analysis made it possible to explore the children's conceptual knowledge about factorization, as well as the influence of semantic context in the encoding of distributive word problems. Similarly, Scheibling-Sève et al. (2022) evaluated the efficacy of a school intervention by looking at the range of solving strategies that children were able to use after the intervention. This focus on solving strategies was also at the core of Gvozdic and Sander (2020) study, who asked first graders to write down the operation they used to solve one-step additive word problems. By analyzing whether children used direct subtraction or indirect addition, they were able to distinguish which problems were solved using

mental simulation, and which were solved using conceptual knowledge about mathematical operations instead. Overall, multiple-strategy word problems thus appear as a promising experimental paradigm to use, together with drawing analysis, to probe the mental representations of adults and children alike.

The issue of cardinality and ordinality

The specific representational differences that we intend to study within this paper relate to a fundamental dimension of mathematics: the distinction between ordinality and cardinality. The concepts of ordinality and cardinality refer to two ontological properties of numbers: ordinality pertains to their position in an ordered sequence (their rank in an ordered list), while cardinality relates to their count value (the number of elements within a set). This distinction is foundational in mathematics (Dantzig, 1945; Frege, 1984; Russell, 1919), particularly in the field of set theory (Dauben, 1990; Suppes, 1972), and research in cognitive psychology has demonstrated that it has implications beyond the realm of formal mathematics.

From a developmental perspective, children's understanding of cardinality (i.e., knowing that counting 1-2-3-4 means that there are four entities) has been investigated in several experimental studies suggesting that children learn to grasp the cardinal meaning of numbers over the first few years of life (Bermejo, 1996; Le Corre & Carey, 2007; Sarnecka & Lee, 2009; Wynn, 1992). This understanding is said to be a crucial step in the development of their mathematical cognition (Geary, 2018; Geary et al., 2018; Shusterman et al., 2016). On the other hand, research on the understanding of the ordinal meaning of numbers, seems to point towards a later development of the ability to use ordinal labels (i.e., "first," "second," "third," and so on) around the fourth and fifth years (Fischer & Beckey, 1990; Hund et al., 2021; Miller, 2015, 2000). Despite receiving less attention than research on cardinality (Goffin & Ansari, 2016), research on ordinality has also shown that the understanding of ordinal position was predictive of 5- and 6-year-olds' arithmetic performance (Cheung & Lourenco, 2019).

The comparative development of these two sides of numerosity has been under scrutiny in recent years, with studies suggesting that the cardinal principle is acquired before children are able to use ordinal labels (Baccaglioni-Frank et al., 2020; Colomé & Noël, 2012; Meyer et al., 2016; Wasner et al., 2015). Overall, these studies highlight that cardinality and ordinality are two crucial notions with distinct developmental trajectories, and that understanding how to use the cardinal and ordinal meanings of numbers is an important part of mathematical development.

However, the influence played by the distinction between cardinality and ordinality among older children and adults

engaged in mathematical reasoning of a higher level has received very little attention in the field in the past decades (Gamo et al., 2010; Verschaffel et al., 1999). Yet, a recent study suggested that adults' mental representation of numerical situations typically fall into one of two categories, on a cardinality-ordinality continuum (Gros et al., 2021). We propose to further investigate this hypothesis by analyzing both drawing productions and strategy choice among children and adults tasked with solving arithmetic word problems.

Encoding differences

Building upon early works on the solving of ordinal problems (Verschaffel et al., 1999) as well as on the difference between age problems and collection problems (Gamo et al., 2010), it was proposed that problems sharing the same mathematical structure could lead either to a cardinal representation or to an ordinal representation, depending on the type of entities being counted (Gros et al. 2021). Due to the non-mathematical knowledge associated with specific entities, the simple mention of daily-life quantities could be enough to tip the scale in favor of one of two possible representations of the same situation (Gros et al., 2020b). For instance, consider the following collection problem:

Paul has five red marbles.
He also has blue marbles.
In total, Paul has 11 marbles.
Jolene has as many blue marbles as Paul, and some green marbles.
She has two green marbles less than Paul has red marbles.
How many marbles does Jolene have?

The hypothesis for this problem was that a problem statement involving counting collections of marbles, which have no inherent order, would highlight the cardinal nature of the numbers used (Gros et al., 2021). Since there is no need to arrange the marbles in a specific order, participants should tend to view the marbles of different colors as separate, distinct sets that are grouped together to be added (see Fig. 1 for a schematic description of their hypothesized mental representation). Thus, when solving the problem, participants should attempt to determine the total number of marbles Jolene has by calculating the two subsets making up her total marble counts. They should count the number of blue marbles she has and add it to the number of green marbles she has. This is done by using a three-step strategy: $11 - 5 = 6$; $5 - 2 = 3$; $6 + 3 = 9$ (Gros et al., 2021).

On the other hand, consider the following duration problem:

Sofia traveled for 5 hours.
Her trip started during the day.

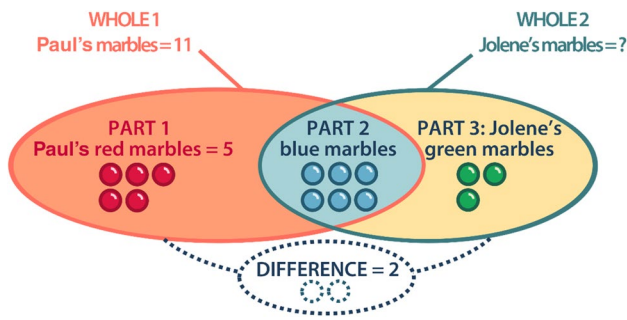


Fig. 1 Example of a hypothesized cardinal representation of the marble problem

Sofia arrived at 11 h.
 Fred left at the same time as Sofia.
 Fred's trip lasted 2 hours less than Sofia's.
 What time was it when Fred arrived?

It was hypothesized that this travel problem, which has the same mathematical structure as the marble problem but involves duration values instead of marble counts, would lead participants to perceive the described situation as ordered along an axis: a timeline of events. This would be due to solvers' knowledge about durations causing them to perceive the different travel times in the problem not as parts and wholes, but as states and transitions along an ordered axis (see Fig. 2 for a schematic description of their hypothesized representation). When considering the problem from this perspective, it becomes easier to understand that there is no need to calculate the duration of Fred's travel nor the hour of his departure. Instead, since Fred and Sofia left at the same time and Fred's travel was 2 h shorter than Sofia's, it follows that Fred arrived 2 h before Sofia. This ordinal representation thus makes it possible for participants to identify a shorter solving strategy: $11 - 2 = 9$, which is seldom used by participants when solving the marble problem (Gros et al., 2021).

Crucially, both the marble problem and the travel problem shared the same mathematical structure (see Fig. 3), and thus both problems could have been solved indifferently

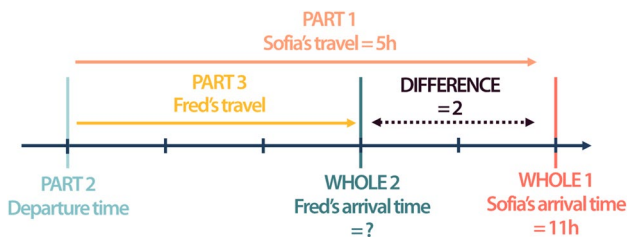


Fig. 2 Example of a hypothesized ordinal representation of the travel problem

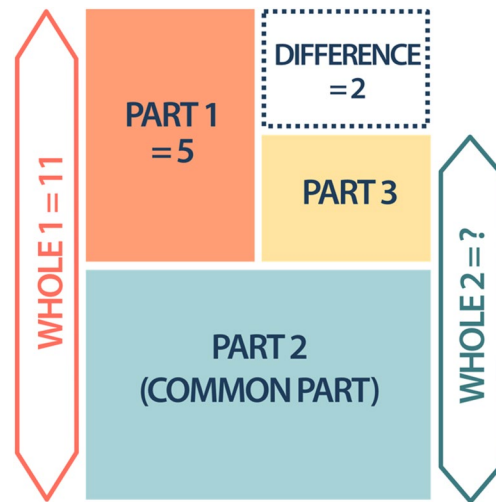


Fig. 3 Deep structure shared by both the marble problem and the duration problem

with both solving strategies. However, data collected among adults showed that participants preferentially use the three-step strategy on cardinal problems, and the one-step strategy on ordinal problems, even when explicitly asked to find the solution involving the shortest number of steps (Gros et al., 2021). Thus, it was proposed that weight problems, price problems, and collection problems would all emphasize the cardinal nature of numbers and lead to a cardinal encoding, due to these quantities usually describing unordered entities. On the other hand, duration problems, height problems, and number of floors problems (problems with an elevator going from one floor to another) would all highlight the ordinal property of numbers instead, due to daily-life knowledge underlining the intrinsic order of the entities they mention (see Gros et al., 2021, for a longer discussion regarding this choice). From this point onward, we will refer to “cardinal problems” as the problems that have statements that mention collections, weights, or prices, and “ordinal problems” as the problems that have statements that mention durations, heights, or elevators.

Interestingly, similar results have also been observed with expert mathematicians who experienced more difficulty using the one-step strategy on cardinal problems than on ordinal problems, even when it was the only available strategy (Gros et al., 2019). This finding illustrates the pervasiveness of encoding effects on one's ability to use a specific solving strategy, and the high cognitive cost of switching from one mental representation to another. This was interpreted as evidence for the influence of non-mathematical information on the semantic encoding of arithmetic word problems, in line with the predictions of the SECO model (Gros et al., 2020b). In the present study, we propose to bring new, converging evidence regarding

the perception of cardinality and ordinality in word problem solving, while at the same time answering a key question: how do drawings relate to mental representations and solving strategies in the arithmetic reasoning of children and adults?

Present study

The current study intends to build upon and contribute to several lines of research. First, we aim to leverage the literature on drawing analysis to propose a systematic analysis of cardinality and ordinality in the mental representations of arithmetic problems. Second, we mean to go one step further than Gros et al.'s (2021) strategy analysis by comparing children's and adults' strategy choices on the same task, using multiple-strategy problems involving cardinal and ordinal quantities. Third, and most importantly, we propose to investigate how the properties of the drawings relate to the hypothesized problem representations and predict participants' solving strategies. By crossing information gathered from strategy choice with insights extracted from drawing production, we expect to get a deeper understanding of the semantic determinants of children and adults' mental representations.

In this perspective, we elected to include both lay adults and fifth grade students in our sample, since both populations possess a certain degree of familiarity with arithmetic word problem solving and additive reasoning. Despite the clear developmental difference between fifth graders and adults, the problems presented in Gros et al. (2021) are within the realm of difficulty for 11-year-olds (Gamo et al., 2010), yet remain challenging to adults (Gros et al., 2019). We asked both groups to consider a series of 12 problems and complete two tasks: solve the problems using as few operations as possible, and make a drawing of the problems that could help someone else understand and solve it. We used the exact same materials and instructions with children and adults to make it possible to compare both groups' productions and strategies.

The rationale was that cardinal problems would elicit a cardinal encoding that would result in drawings presenting features prototypical of a cardinal representation, while also leading participants to preferentially use the three-step solving strategy. On the other hand, we expected that ordinal problems would lead to the production of drawings with features highlighting the ordinality of the numerical values, as well as to a higher rate of one-step solving strategies. Finally, we also made the hypothesis that independently from the mention of cardinal or ordinal quantities in the problem statements, the ordinality rating of the drawings would significantly predict the likelihood of participants using the one-step strategy.

Methods

Participants

Using the BUCSS R package (v1.2.1; Anderson & Kelley, 2018), a minimum sample size of 45 was determined based on results from a previous study using similar materials (Gros et al., 2021 – Experiment 4), after correction for uncertainty and publication bias following Anderson, Kelley, and Maxwell's recommendations (Anderson et al., 2017). We used a high level of targeted statistical power (0.95) to account for the uncertainty linked to the fact that previous studies only analyzed participants' solving strategies, not their drawings' structural features. Participants were recruited from two populations: a group of 59 children in fifth grade, recruited among several schools from the Paris region (27 girls, $M_{age} = 11.00$ years, $SD = 0.36$), and a group of 52 adults from the Paris region (36 women, $M_{age} = 26.86$ years, $SD = 9.72$). All participants spoke French fluently. None had previously participated in any similar experiment.

Materials and procedure

Each participant was presented with a set of 12 different problems: six problems using ordinal quantities ("duration," "height," and "elevator" problems, see Table 1) and six problems using cardinal quantities ("collection," "price," and "weight" problems, see Table 2), as defined in Gros et al. (2021).

Each participant was given a 13-page booklet. On the first page, detailed instructions asked participants to solve a series of math problems using as few operations as possible. It was explained that an operation was defined by two operands, an operator, and a result (e.g., $2 + 2 = 4$), and that they needed to write down any operation they used, even if they resorted to mental calculation. Then, it was indicated that for each problem, they had to make a diagram meant to "help someone understand the problem and solve it." We used this specific instruction to maximize the chances of participants' producing informative diagrams depicting the mathematical relations as they understood them. Indeed, previous works have shown that "drawing to communicate" stimulates pragmatic inferences that help the drawer select the most informative pieces of information to include in the drawings (Fan et al., 2020).

The next 12 pages of the booklets were dedicated to the 12 problems participants were tasked with solving. Each page was divided into four sections: first, the problem statement was presented, next to it was a "draft" space, providing participants with the freedom to work

Table 1 English translation of the cardinal problems used in this study. The numerical values respected the following rule: $z < 4 < x < y < 15$

Quantity used	Problem statement
Weight	A bag of pears weighs x kilograms. It is weighed with a whole cheese. In total, the weighing scale indicates y kilograms. The same cheese is weighed with a milk carton. The milk carton weighs z kilograms less than the bag of pears. How much does the weighing scale indicate?
Weight	Tom takes a Russian dictionary weighing x kilograms. He also takes a Spanish dictionary. In total, he is carrying y kilograms of books. Lucy takes Tom's Spanish dictionary and a German dictionary. The German dictionary weighs z kilograms less than the Russian dictionary. How many kilograms is Lucy carrying now?
Price	In the first meal on the menu, there is a chocolate cake costing x euros. The meal also includes an omelet with mushrooms. In total, the first meal costs y euros. In the second meal on the menu, there is the same mushroom omelette, and an apple pie. The apple pie costs z euros less than chocolate cake. How much does the second meal cost?
Price	In the stationery shop, Antony wants to buy a x -euro ruler. He also wants a notebook. In total, that will cost him y euros. Julie wants to buy the same notebook as Antony, and an eraser. The eraser costs z euros less than the ruler. How much will Julie have to pay?
Collection	Paul has x red marbles. He also has blue marbles. In total, Paul has y marbles. Jolene has as many blue marbles as Paul, and some green marbles. She has z green marbles less than Paul has red marbles. How many marbles does Jolene have?
Collection	Sarah owns x goldfish. Her other pets are all iguanas. In total, she owns y pets. Bobby is pet-sitting Sarah's iguanas during the holidays, he puts them with the turtles he owns. Bobby owns z turtles less than Sarah owns goldfish. How many pets are there at Bobby's during the holidays?

through the problem in their preferred manner. Below, there was a "calculations and result" space for recording their problem-solving strategies. Finally, the bottom half of the page was reserved for a "diagram" space where participants were invited to create a schematic drawing representing the situation described in the problem. The order of problems within the booklets was arranged in a pseudorandom sequence, with cardinal and ordinal problems alternating on each page. This sequencing was designed to enhance the likelihood of participants recognizing the underlying deep structure shared across all the problems. Additionally, two versions of the booklets were created, each with an inverse problem order, to control for potential primacy effects.

Drawing scales

Participants' diagrams were analyzed using two custom scales, designed to evaluate to what extent they featured either ordinal or cardinal characteristics. The scales were created around the idea that the hypothesized semantic encoding of the problems should permeate the participants' drawings and lead to specific features occurring in problems sharing a similar encoding. The scales were created prior to any collection of data, based on the ontological properties of cardinal and ordinal representations, as defined in Gros et al. (2021). Each scale included four criteria.

The cardinal scale included four items that were deemed typical features of a cardinal representation of the problems

Table 2 English translation of the ordinal problems used in this study. The numerical values respected the following rule: $z < 4 < x < y < 15$

Quantity used	Problem statement
Duration	<p>The construction of the cathedral took x years. Before constructing it, the plans had to be made. The construction of the cathedral was completed in year y. The construction of the castle started at the same time as the construction of the cathedral. The construction of the castle took z years less than the construction of the cathedral. When was the construction of the castle completed?</p>
Duration	<p>Sofia traveled for x hours. Her trip started during the day. Sofia arrived at y. Fred left at the same time as Sofia. Fred's trip lasted z hours less than Sofia's. What time was it when Fred arrived?</p>
Height	<p>Slouchy Smurf is x-centimeter tall. He climbs on a table. Now he reaches y centimeters. Grouchy Smurf climbs on the same table as Slouchy Smurf. Grouchy Smurf is z centimeters shorter than Slouchy Smurf. What height does Grouchy Smurf reach when he climbs on the table?</p>
Height	<p>Obelix's statue is x-meter tall. It is placed on a pedestal. Once on the pedestal, it reaches y meters. Asterix's statue is placed on the same pedestal as Obelix's. Asterix's statue is z meters shorter than Obelix's. What height does Asterix's statue reach when placed on the pedestal?</p>
Elevator	<p>Naomi takes the elevator and goes up x floors. She left from the floor where her grandparents live. She arrives at the y th floor. Her brother Derek also takes the elevator from their grandparents' floor. He goes up z floors less than Naomi. At what floor does Derek arrive?</p>
Elevator	<p>Karen takes the elevator and goes up x floors. She left from the floor where the gym is. She arrives at the y th floor. Yohan also takes the elevator from the floor where the gym is. He goes up z floors less than Karen. At what floor does Yohan arrive?</p>

Criteria	Examples	Rating
(a) Are there clusters of identical elements?		Present <input type="checkbox"/> Absent <input type="checkbox"/>
(b) Are there sets containing more than one element?		Present <input type="checkbox"/> Absent <input type="checkbox"/>
(c) Are there embedded sets?		Present <input type="checkbox"/> Absent <input type="checkbox"/>
(d) Are there one-to-one correspondences between an element and a value?		Present <input type="checkbox"/> Absent <input type="checkbox"/>

Fig. 4 Cardinal drawing scale provided to the independent raters to score the drawings

(see Fig. 4). Since cardinality refers to the total number of elements within a set, regardless of their order, we expected cardinal representations to include some or all of the following prototypical features: (a) presence of identifiable clusters of elements, each expressing a specific value, (b) drawings of sets containing several entities, indicative of a group-based representation, (c) inclusion of sets within each other, to account for the part/whole relationships highlighted by a cardinal encoding, and (d) explicit correspondence between a specific entity and an assigned value. As depicted in Fig. 4, we proposed several examples for each of these categories in the scoring sheet, so that the raters in charge of scoring the drawings would be able to identify the hypothesized features without being informed of our hypotheses.

Following the same procedure, the ordinal scale included four items involving the most relevant features of what we hypothesized to be a prototypically ordinal representation (see Fig. 5). Considering the properties of ordinality, we predicted that an ordinal representation would include some of the following prototypical features: (a) presence of axes on which to place different problem values, (b) presence of graduations quantifying the numerical distance between two values on an axis, (c) side-by-side presentation of different axes indicative of value comparisons, and (d) presence of intervals denoting a numerical distance between two other values.

Importantly, every item in both scales could have been used to describe any of the problems in the experiment. For instance, it is entirely possible to make a diagram of the travel problem using only the cardinal features described in the cardinal scale, and it is equally possible to make a diagram of the marble problem using only the features of the ordinal scale. Yet, we made the hypothesis that features in the cardinal scale would predominantly appear on problems meant to elicit a cardinal encoding, whereas features in the

ordinal scale would predominantly appear on the ordinal scale instead.

Double-blind scoring was performed by two independent raters who were unaware of the hypotheses being tested. The two drawing scales were introduced to them, and they were given the opportunity to ask questions about their different criteria and examples. Once they were confident they understood the meaning of each of the eight items, they were asked to score the entirety of the diagrams produced by the participants in both groups. Thus, both rates scored each of the 1,332 diagrams on the eight items. After initial independent scoring, the two raters reached perfect agreement on 91.01% of cases. The occasional discrepancies were due to slight differences in the interpretation of some of the scales' criteria. Notably, one rater was somehow more liberal in what constituted an interval during the initial coding. Yet, after discussion both agreed to a common definition. Another topic of discussion was raised when one rater judged it impossible to have graduations even without an axis, while the other considered some diagrams to feature graduations without any drawn axis. They decided to settle on the latter perspective. Cohen's Kappa coefficient for interrater reliability was calculated to determine consistency among raters based on their initial rating. The result ($\kappa = .726, SE = 0.012$) expressed substantial agreement between raters, according to Landis and Koch's typology (Landis & Koch, 1977). After discussion between themselves and without the authors, the raters managed to decide on common rules for ambiguous cases. They reached 100% agreement on their second round of scoring. Based on the raters' final assessment of which criteria were met by each drawing, two scores were calculated: a cardinal drawing score, and an ordinal drawing score. The cardinal score (from 0 to 4) indicated how many of the four cardinal criteria were met by the drawings (see Fig. 4). The ordinal score (from 0 to 4)

Criteria	Examples	Rating
(a) Are there axes?		Present <input type="checkbox"/> Absent <input type="checkbox"/>
(b) Are there graduations?		Present <input type="checkbox"/> Absent <input type="checkbox"/>
(c) Are there axes side by side?		Present <input type="checkbox"/> Absent <input type="checkbox"/>
(d) Are there intervals?		Present <input type="checkbox"/> Absent <input type="checkbox"/>

Fig. 5 Ordinal drawing scale provided to the independent raters to score the drawings

indicated how many of the four ordinal criteria were met by the drawings (see Fig. 5).

Strategy choice interpretation

Participants’ solving strategies were extracted from their self-report of the operations they had performed. A problem was considered as correctly solved when the correct result came with the appropriate calculations. The strategies leading to success were categorized either as a one-step strategy or as a three-step strategy. When the written operations were correct and the written solution was within ± 1 of the correct result, this was deemed a calculation error and problems were still considered as correctly solved.

When participants wrote down operations that did not provide the answer to the problem, or when they simply provided an erroneous answer, their response was labeled “Error.” Because participants were instructed to write down every operation they performed, and because numerical values were chosen so that they could not lead to two identical values being calculated using two distinct strategies (i.e. $x + y \neq z - y$ and so on), it was always possible to trace back the strategies used by participants as long as they wrote down the solution and at least one of the operations performed. If participants forgot to report one of the first two operations of the three-step strategy (i.e., they forgot either “Part 1 – Difference = Part 3” or “Whole 1 – Part 1 = Part 2”) but still reported the result of said operation and used it to perform the final operation of the three-step strategy (i.e. “Part 3 + Part 2 = Whole 2”), then the response was labeled as a three-step strategy. The rare cases in which the solution

was given with no explanation were considered as incorrect (which occurred in less than 1% of the trials).

Results

Diagram analysis

Both children and adults engaged with the drawing task by producing different diagrams displaying varied characteristics. Sample diagrams of cardinal problems by children and adults are attached in the supplemental materials (Fig. A). Similarly, examples of children and adults’ productions for ordinal problems can be seen in the Online Supplemental Materials (OSM; Fig. B). Most diagrams included some schematic elements, as well as numerical values and text labels (see diagrams reported in OSM Fig. A and Fig. B). Some diagrams were particularly depictive with, for instance, adults drawing each individual turtle at Bobby’s house (see second diagram in OSM Fig. A). Others were mostly abstract, with no illustrative element other than the schematic representation of the relations between the numerical values (see second diagram in OSM Fig. B).

We analyzed the diagrams produced for each problem by each participant, using the 4-point cardinal scale and the 4-point ordinal scale. Overall, there was no major surprise in the diagrams, and while the rules behind the rating scales had to be further specified by the raters to decide on ambiguous drawings, the two rating scales captured a wide variety of cases, each item appearing on average in 14.40% of the drawings. The two criteria that were most often found in the participants’ diagrams were “presence of one-to-one

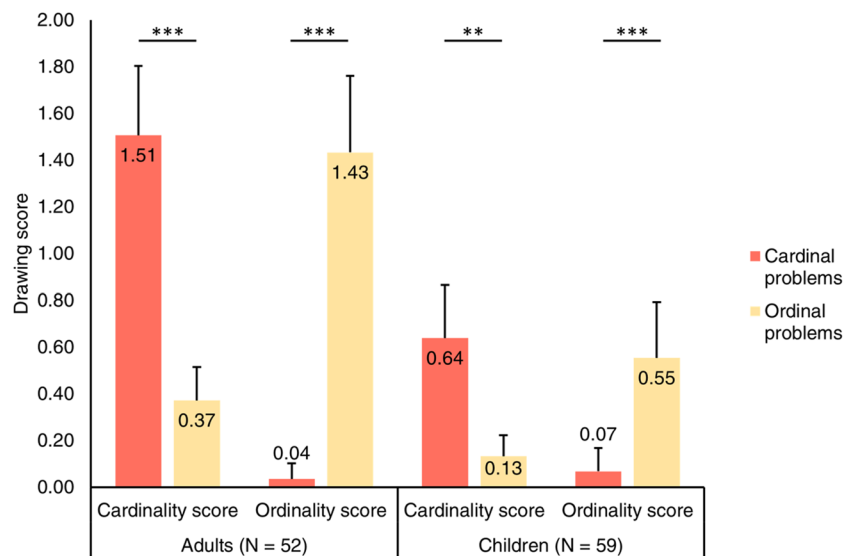


Fig. 6 Children’s and adults’ mean cardinal and ordinal scores depending on the type of problems (problems with cardinal quantities versus problems with ordinal entities). Vertical bars denote 0.95 confidence intervals. *** $p < .001$, ** $p < .01$ Tukey-adjusted least squares means

correspondences” (32.58% of the drawings) and “presence of axes” (19.97% of the drawings). The two least used items in the scale were “presence of embedded sets” (3.00% of the drawings) and “presence of intervals” (6.76% of the drawings). Figure 6 details the mean cardinal and ordinal scores of the drawings depending on the type of quantity used in the problems, for each population. For each rating scale, we compared the scores attributed to participants’ drawings on cardinal and on ordinal problems.

Ordinal score analysis

Adults’ drawings of ordinal problems had a higher ordinal score on average ($M = 1.43$, $SD = 1.20$) than their drawings of cardinal problems ($M = 0.04$, $SD = 0.24$). Similarly, children’s drawings of ordinal problems were also rated higher on the ordinal scale on average ($M = 0.55$, $SD = 0.93$) than their drawings of cardinal problems ($M = 0.07$, $SD = 0.39$). To evaluate the statistical significance of this difference, we used a linear mixed model with the ordinal score as the dependent measure. We included experimental group (children vs. adults) and type of problem (cardinal problem vs. ordinal problem) as fixed effects, and we added participants and problem statements as random effects to account for the repeated measures in the experimental design. As predicted, there was a significant effect of problem type on the ordinal score attributed to the drawings ($F = 109.86$, $p < .001$; Type III SS ANOVA). Pairwise comparisons using Tukey-adjusted LS-means revealed that this difference was significant for children ($t(13.4) = 5.19$, $p < .001$) as well as for adults ($t(14.4) = 14.64$, $p < .001$). Interestingly, there was also a main effect of the group of participants (children vs.

adults) ($F = 24.58$, $p < .001$; Type III SS ANOVA). Indeed, children’s drawings were rated lower on average on the ordinal scale ($M = 0.31$, $SD = 0.76$) than adults’ drawings ($M = 0.73$, $SD = 1.11$). Finally, there was a significant interaction between the experimental group (children vs. adults) and the type of problems (cardinal vs. ordinal) on the ordinal score ($F = 153.60$, $p < .001$; Type III SS ANOVA). This interaction may be due to the fact that children’s diagrams tended to be less complex than adults’, which led to them scoring lower than adults on the ordinality score for ordinal problems, but not for cardinal problems, due to a floor effect on cardinal problems.

To better understand the extent to which each of the four criteria associated with the ordinal scale accounted for the observed differences between cardinal and ordinal problems, we also looked at the role played by each item in the overall score. Figure 7 details how often each of the four items of the ordinal scale was present in a participant’s drawing. Interestingly, the item that was most often found on the drawings of ordinal problems was item (a) (presence of axes), which appeared in 58.65% of adults’ drawings of ordinal problems and in 20.90% of children’s drawings of ordinal problems. Conversely, the least observed item was item (d) (presence of intervals), which was only present in 16.67% of adults’ drawings of ordinal problems and 5.65% of children’s drawings of the same problems.

We created generalized mixed models using each criterion as a binary outcome, group and type of problem as fixed effects, and participants as well as problem statements as random effects to evaluate whether each criterion appeared more frequently on ordinal than on cardinal problems. Results confirmed that the nature of the problems (ordinal

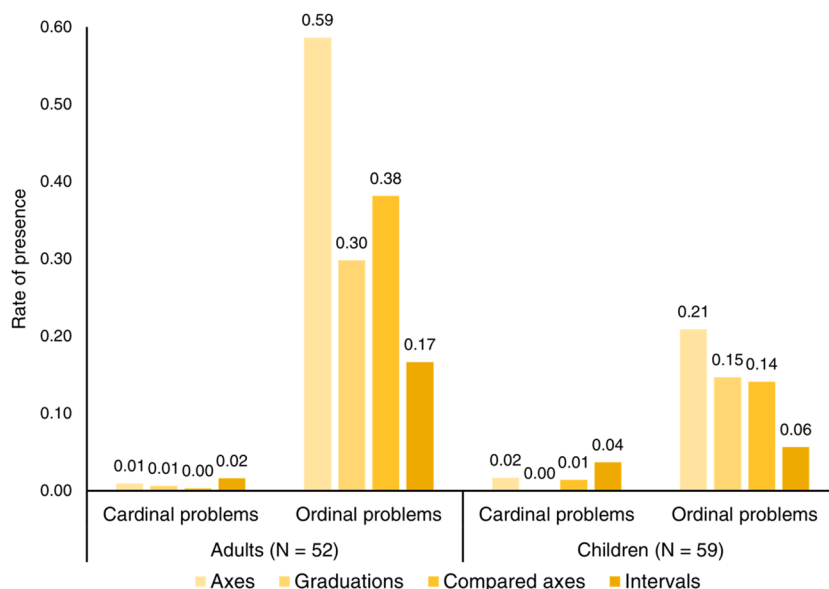


Fig. 7 Rate of presence in participants’ drawings of each of the ordinal scale’s four criteria

vs. cardinal) had a significant effect on the presence of the four ordinal criteria, be it (a) axes ($Z = 8.79, p < .001$), (b) graduations ($Z = 4.33, p < .001$), (c) compared axes ($Z = 5.70, p < .001$) or (d) intervals ($Z = 5.91, p < .001$).

For further analysis of the presence of each cardinal criterion for the three cardinal quantities and the three ordinal quantities, see Fig. C in the OSM. The distribution of the four ordinal criteria were mostly stable across the three ordinal quantities, with the exception of the “presence of graduations” criterion, which was more often met on drawings of elevator problems (47.30% of cases) than on drawings of duration problems (12.61% of cases) or of height problems (5.41% of cases). This discrepancy may be due to the fact that in elevator problems, each unit (and thus each graduation) corresponds to a distinct, easy to draw entity (each floor on the building). Drawing the floors and drawing the graduations are relatively similar processes, so one might have helped the other, which would explain the high rate of graduations in diagrams of elevator problems. Apart from this criterion, there was no noticeably unusual pattern in the distribution of criteria across quantities. The most reliable criterion to assess ordinality appears to be “presence of axes,” since it appeared in 36.94% of drawings of elevator problems, in 31.98% of duration problems, and in 46.85% of height problems, while it was hardly present in drawings of cardinal problems (1.35% overall).

Cardinal score analysis

Regarding the cardinal score attributed to participants drawings, we had made the hypothesis that participants’ drawings would display more typically cardinal features on cardinal problems than on ordinal problems. Indeed, adults’ drawings

of cardinal problems had a higher cardinal score on average ($M = 1.51, SD = 1.09$) than their drawings of ordinal problems ($M = 0.37, SD = 0.52$). Similarly, children’s drawings of cardinal problems were also rated higher on average on the cardinal scale ($M = 0.64, SD = 0.89$) than their drawings of ordinal problems ($M = 0.13, SD = 0.35$). We used a linear mixed model to evaluate the factors influencing the cardinal score, with experimental group (children vs. adults) and type of problem (cardinal problem vs. ordinal problem) as fixed effects, and participants and problem statements as random effects. As hypothesized, there was a significant effect of problem type on the drawings’ cardinal score ($F = 46.25, p < .001$; Type III SS ANOVA). Pairwise comparisons using Tukey-adjusted LS-means revealed that this difference was significant for children ($t(11.8) = 4.13, p < .01$) as well as for adults ($t(12.3) = 9.16, p < .001$). Additionally, there was a main effect of the group of participants (children vs. adults) ($F = 57.14, p < .001$; Type III SS ANOVA). Indeed, as with the ordinal score, children’s drawings tended to score lower on average on the cardinal scale ($M = 0.39, SD = 0.72$), compared to adults’ productions ($M = 0.94, SD = 1.03$). Finally, there was a significant interaction between the experimental group (children vs. adults) and the type of problems (cardinal vs. ordinal) on the cardinal score as well ($F = 74.07, p < .001$; Type III SS ANOVA). As with the ordinal score, we attributed this interaction to the lower complexity of children’s drawings, impacting the cardinal score of cardinal problems more heavily than the cardinal score of ordinal problems, due to a floor effect on the latter.

We also computed the proportion of each of the four criteria by problem type in participants’ drawings (see Fig. 8). Interestingly, the role played by each criterion appeared to be less homogeneous than for the ordinal scale. Namely, while

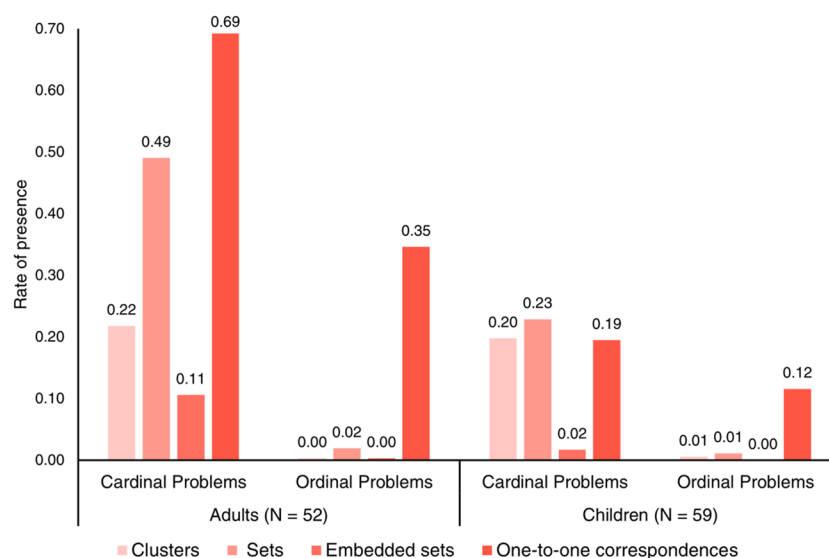


Fig. 8 Rate of presence in participants’ drawings of each of the four criteria on the cardinal scale

the “clusters” and “sets” criteria tended to follow a similar pattern of moderate frequency on cardinal problems (20.72% and 35.14%, respectively) and nil-absence on ordinal problems (0.45% and 1.50%, respectively), the two remaining criteria displayed a different behavior. The “embedded sets” criterion was only identified in a limited number of few participants’ productions. Even on cardinal problems, adults’ drawings displayed embedded sets in only 10.58% of cases, while children used them in just 1.69% of cardinal problems. This low score may be attributed to the fact that this criterion is dependent upon another one (there needs to be “sets” in order to have “*embedded sets*”). Finally, the “one-to-one correspondence” criterion appeared to be less typical of cardinal drawings than we expected. Indeed, while the criterion was met in by a majority of adults’ drawings of cardinal problems (69.23%), their drawings of ordinal problems also included this feature in 19.49% of cases, as it was sometimes used to assign numerical values to graduations labels.

To evaluate the impact of the type of problem on each cardinal criterion, we created generalized mixed models using each criterion as a binary outcome. We included type of problem and experimental group as fixed effects, and we used participants and problem statements as random effects to account for the repeated measures in the experiment. Results showed that, regardless of the aforementioned variations between items, each individual criterion in the cardinal scale was significantly more present on drawings of cardinal problems than on drawings of ordinal problems, be they (a) clusters of identical elements ($Z = 2.80, p < .01$), (b) sets ($Z = 8.23, p < .001$), (c) embedded sets ($Z = 3.75, p < .001$), or (d) one-to-one correspondences ($Z = 6.26, p < .001$).

For further analysis of the presence of each ordinal criterion for the three cardinal quantities and the three ordinal quantities in the problems, see Fig. D in the OSM.

As with the ordinal scale, the distributions of the four cardinal criteria were mostly stable across the three cardinal quantities, with the exception of the criterion “presence of clusters of identical elements.” Indeed, this criterion was often met on drawings of collection problems (48.65%), but more rarely on drawings of weight problems (10.81%) or of price problems (2.70%). This may be due to the fact that drawing each marble in the problem, for instance, comes more naturally than drawing each kilogram corresponding to the weight of a wheel of cheese. Except for this item, there was no unexpected pattern in the distribution of cardinal criteria across quantities. The “presence of sets” criterion, however, seemed to be a good measure of cardinality across all quantities, since it was present in 43.69% of collections problems, in 31.53% of weight problems, and in 30.18% of height problems, while almost never featured in drawings of ordinal problems (0.45% overall).

Thus, adults and fifth graders alike were more likely to use ordinal features (axes, graduations, etc.) than cardinal

features (clusters, sets, etc.) on ordinal problems, and conversely on cardinal problems. In sum, drawing analysis shows that in both populations, the mention of ordinal (or respectively cardinal) quantities in the problem statements seems to result in representations featuring a higher number of ordinal (respectively cardinal) features. For a quantity-by-quantity comparison of the cardinal and ordinal scores, see Fig. E in the OSM.

Strategy analysis

Second, we had made the prediction that problems with ordinal quantities would facilitate the use of the one-step strategy compared to problems with cardinal quantities. Across both groups, we evaluated whether participants did use the one-step strategy more often on problems involving ordinal quantities than on problems involving cardinal quantities. Figure 9 details the participants’ use of each strategy depending on the type of quantity featured in the problems.

We used a generalized mixed model with the use of the one-step strategy as its binary outcome to evaluate the factors influencing participants’ strategies. We included the type of problem (cardinal vs. ordinal) and the experimental group (children vs. adults) as fixed effects, and we used participants and problem statements as random effects accounting for the repeated measures of the design. Results revealed a main effect of problem type on the rate of use of the one-step strategy ($Z = 4.25, p < .001$), with a higher rate for ordinal problems ($M = 43.99%$) than for cardinal problems ($M = 16.97%$), as hypothesized. There was also a main effect of the experimental group ($Z = 4.57, p < .001$), since adults were more likely than children to find the one-step solving strategy.

Finally, there was an interaction between problem type and experimental group ($Z = 3.75, p < .001$), which suggests that fifth graders’ use of the one-step solving strategy was more impacted by the cardinal versus ordinal nature of the problems than that of adults. Indeed, while their use of the one-step strategy was relatively high on ordinal problems ($M = 38.98%$), their performance substantially dropped on cardinal problems ($M = 7.63%$). LS pairwise comparisons computed using the Emmeans R package (v1.8.9; Lenth et al., 2023) showed that this difference was statistically significant ($z\text{-ratio} = 7.16, p < .001$). On the other hand, the difference between the two types of problems was less important for adults, although it remained significant. Indeed, their mean rate of use of the one-step strategy was higher on ordinal problems ($M = 49.68%$) than on cardinal problems ($M = 27.56%$); $z\text{-ratio} = 4.25, p < .001$. Thus, our hypothesis was confirmed among both populations, despite adults performing better than children, especially on cardinal problems.

In addition, we also looked at the distribution of three-step strategies, to investigate whether its use was also linked

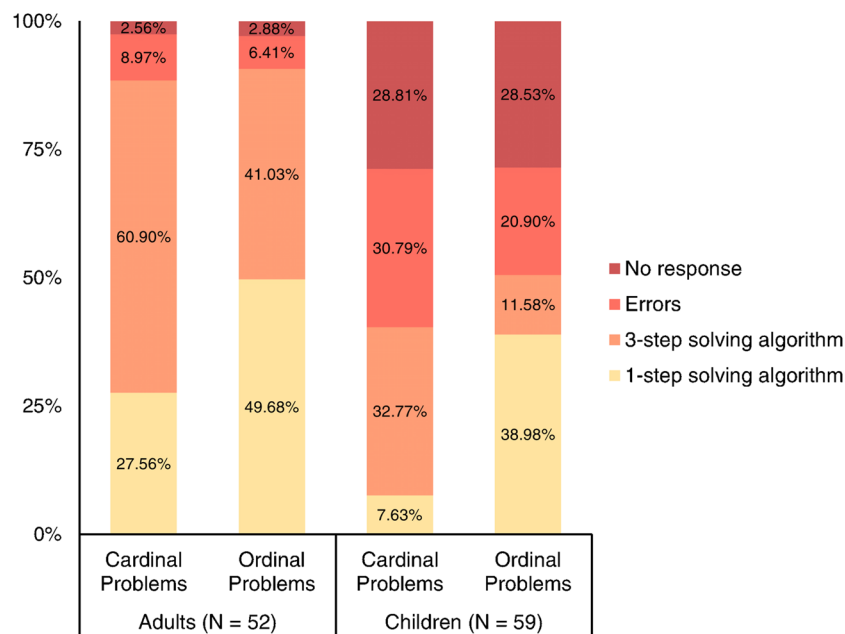


Fig. 9 Children’s and adults’ mean rate of use of the two solving strategies depending on the type of quantities used in the problems

to the cardinal versus ordinal nature of the problems. Using a generalized mixed model similar to the one for the one-step strategy, we looked at the effects of problem type and experimental group on participants’ responses. Again, we found a main effect of problem type ($Z = 3.27, p < .01$), a main effect of experimental group ($Z = 4.38, p < .001$), and an interaction between the two ($Z = 2.34, p < .05$). Least-squares pairwise comparisons calculated with Emmeans (v1.8.9; Lenth et al., 2023) showed that among children, the three-step strategy was more often used on cardinal problems ($M = 46.03\%$) than on ordinal problems ($M = 16.21\%$); z -ratio = 4.89, $p < .001$. This was also the case for adults: the use of the three-step strategy was more frequent on cardinal problems ($M = 62.50\%$) than on ordinal problems (42.40%); z -ratio = 3.27, $p < .01$. Hence, the choice of a solving strategy in both groups is influenced by the cardinal versus ordinal nature of the problem’s quantities.

Finally, we also looked at participants’ preferred strategies to solve the problems depending on the quantities they involved. Figure G in the OSM details the rate of use of each strategy (as well as the rate of errors and absence of response) by adults and children, for the three ordinal and the three cardinal quantities. The distribution pattern tended to be homogeneous across cardinal problems, as well as across ordinal problems, with one exception. Indeed, the height problems were the only problems where adults’ rate of use of the one-step solving strategy (30.77%) appeared to be descriptively lower than that of children (34.76%). While this difference was not statistically significant (z -ratio = 0.61, $p = .99$), it was nevertheless surprising, as it seems to

suggest that, as adults progress in their ability to use longer, multiple-step strategies (as evidenced by the increase in three-step strategy use between fifth grade and adulthood), this progression may sometimes be detrimental to their ability to identify shorter, more effective solving strategies (in this case, the one-step strategy). Future work on this issue may yield interesting insights into the developmental trajectory of strategy use and height representations.

Analysis of the links between drawings and strategies

Third, we looked at how the ordinality of the drawings predicted which problems would be solved using the one-step strategy, independently from the cardinal versus ordinal nature of the problems themselves. Because the number of successes and failures varied between participants, we used a generalized linear mixed model with a binomial distribution to evaluate the extent to which the ordinal drawing score predicted participants’ propensity to successfully use the one-step strategy to solve the problems. The cardinal versus ordinal nature of the drawings was used as a fixed effect, as was the ordinal drawing score. We used participants as a random effect, as well as problem statements, to account for the design’s repeated measures (each participant had to attempt to solve 12 problems). Analysis of the model showed that there was a significant effect of the ordinal score of the drawings on participants’ rate of use of the one-step strategy, even after accounting for the influence of the type of problem ($Z = 2.34, p < .05$). In other words, in accordance with our hypothesis,

a higher ratio of ordinal features in the drawings predicted higher chances to use the one-step strategy, regardless of whether participants were attempting to solve a cardinal or an ordinal problem. This effect was not present, however, when using the rate of the three-step strategy as the dependent variable of the model ($Z = 0.71, p = .48$). Interestingly, the cardinal scale showed the reverse pattern: a generalized mixed model revealed that the cardinality of the drawings had a main effect on participants' likelihood of using the three-step solving strategy to solve the problems ($Z = 2.76, p < .01$), but not on their propensity to use the one-step strategy ($Z = 0.31, p = .75$). In other words, the ordinality of the drawings predicted participants' propensity to find the one-step solving strategy, while the cardinality of the productions predicted the likelihood of participants using the three-step solving strategy instead, regardless of the type of problems to be solved.

Discussion

By resorting to a drawing task coupled with a solving task, we were able to gain new insights into children's and adults' representations of arithmetic word problems. The drawing analysis allowed us to get a new, sharper look at participants' conceptualization of cardinal and ordinal problems without resorting to explicit verbalization. Drawing analysis supported the idea that cardinal problems foster representations involving sets and unordered collections, whereas ordinal problems favor the construction of representations involving axes, graduations, and intervals. This was the case for both children and adults, which speaks volumes for the pervasiveness of the influence of non-mathematical knowledge on mathematical reasoning across the lifespan. Indeed, despite children's drawings scoring lower, on average, than adults' drawings on both scales (which we attributed to less detailed and more pictorial drawings) the difference between the two scores remained unequivocally significant in children's productions on both cardinal and ordinal problems.

Interestingly, the use of these two drawing scales also made it possible to compare competing theories regarding the representation of arithmetic word problems. Indeed, as mentioned in the *Introduction*, the SECO model (Gros et al., 2020b) predicts that solvers encode an interpreted structure depending on the mathematical and world semantics evoked by the problem statement (in our case, a representation involving either mostly cardinal features, or mostly ordinal features). However, we believe that the two competing theories that the SECO model challenges would make different predictions regarding participants' drawings in our experiment. According to the Schema model (Kintsch & Greeno, 1985), learners use the problem statement to extract a propositional structure that will be used to identify and implement a solving schema with the relevant numerical information in

the problems. This schema then triggers a calculational strategy to find the solution (Kintsch & Greeno, 1985). In other words, according to this theory, the propositional structure determines the type of representation being constructed, as well as the solving strategies used. In this view, participants' representations of the numerical elements in the cardinal and ordinal problems should have been linked to the propositional wording of the problem statements, rather than to the entities mentioned in the problems. For instance, when describing the difference between Part 1 and Part 3, the marble problem used the wording "She has *two green marbles less than* Paul has red marbles" and the travel problem "Fred's trip lasted *2 hours less than* Sofia's." According to Kintsch and Greeno's (1985) theory, both sentences include a HAVE-LESS-THAN proposition, which should cue the use of the same DIFFERENCE schema. Thus, under the framework of the schema theory, participants' mental representation of this difference should be similar in both instances, which would result in similar diagrams being used to represent it in their drawings. Yet, the drawing analysis revealed that, contrary to this prediction, this difference tended to be represented by an interval in ordinal problems, and by a set in cardinal problems (see Gros et al., 2020b, for an in-depth discussion of the competing predictions between the schema theory and SECO).

Similarly, the Situation Problem Solver model – the other competing theory – states that, when reading a problem statement, learners construct an episodic situation model specific to the problem (Reusser, 1990). This situation model is expected to include every functional relation described in the problem statement (Johnson-Laird, 2010). This approach thus proposes that every problem gives rise to a particular interpretation, an idiosyncratic representation. On the other hand, SECO suggests that structural regularities emerge in the representation of different problems, due to underlying semantic dimensions driving their encoding. In terms of problem drawings, the situation model approach would certainly predict that different problems elicit different drawings, but it would not predict a consistent difference between the drawings of cardinal and ordinal problems. SECO on the other hand,

predicts that individuals' representation of a numerical situation tends to highlight either the cardinality of its numbers or their ordinality, depending on the world semantics attached to the problem statement (Gros et al., 2020b, 2021). The drawing analysis showed that diagrams of cardinal problems did share common structural features that were different from the structural features shared by the diagrams of ordinal problems.

Additionally, SECO also proposes that individuals' representation of a numerical situation tends to focus either on the cardinality of its numbers or on its ordinality, but rarely on both (Gros et al., 2021). By looking at the distribution of

cardinal and ordinal scores, we were able to assess this prediction in our experiment (see Fig. F in the OSM). Interestingly, among all the recorded answers, no drawing received a score equal to or greater than 2 on both scales simultaneously. In fact, whenever a drawing's cardinal score exceeded 2, its ordinal score was consistently zero. Conversely, if the ordinal score was 2 or higher, the corresponding cardinal score was always less than or equal to 1. This is compatible with the idea that individuals' representations tend to include either cardinal or ordinal features, but not both simultaneously.

As for strategy choice, Gros et al.'s (2021) findings were replicated in this experiment, since participants in both groups tended to use the one-step strategy on ordinal problems whereas they preferentially used the three-step strategy on cardinal problems, regardless of the instructions asking them to solve the problems using as few operations as possible. The fact that those effects could be highlighted with fifth graders as well as with adults on the same problems speaks to the universality of such encoding constraints. From beginner problem solvers to adults with years of experience performing small additions and subtractions, the difference between cardinal and ordinal problems remained influential.

Finally, and most importantly, the apparent ordinality of the drawings made by children and adults was predictive of their propensity to use the one-step strategy to solve the problems – regardless of whether the problems were cardinal or ordinal to begin with. This suggests that the drawings' features that we were attentive to were indeed relevant to explain participants' reasoning process and strategy choice. In other words, a drawing featuring axes instead of sets was more likely to indicate that children would find the one-step strategy, regardless of whether the problem involved cardinal or ordinal quantities to begin with. From an educational perspective, this is especially interesting, since it suggests that the presence of specific features within a student's drawing can reveal the inadequacy of their mental representation of the problem and predict their inability to successfully complete a task (in this case, finding the shortest solving strategy). Thus, analyzing students' drawing productions may open the way for targeted interventions aiming at fostering a semantic recoding of a suboptimal problem representation.

In addition to the insights gained from studying learners' drawings, the use of a drawing task also made it possible to assess the robustness of the semantic congruence effects previously described. Indeed, the literature suggests that drawing production has several remarkable benefits: drawing tasks have been shown to improve encoding and recall (Draschkow et al., 2014; Meade et al., 2018; Roberts & Wammes, 2021; Van Meter & Garner, 2005; Wammes et al., 2016, 2017, 2018), while also helping with understanding complex notions (Schmeck et al., 2014), engaging in scientific thinking (Fan, 2015), and learning STEM content

(Wu & Rau, 2019). Regarding problem solving, research indicates that drawing can help students draw crucial inferences, revealing information that was only implicit in the problem statement (Cox, 1999; Larkin & Simon, 1987). In interventions designed to improve mathematical word problem solving, drawing practice has even been shown to increase children's progress (Csíkos et al., 2012; Sharp & Shih Dennis, 2017; Van Essen & Hamaker, 1990). In this perspective, asking participants to make a drawing of the problems they are attempting to solve could have resulted in them gaining a finer understanding of the problems' mathematical structure, thus increasing the likelihood that they would manage to identify the isomorphism between all the problems, and manage to use the one-step strategy on every problem, regardless of the quantities they featured. The fact that the addition of a drawing task did not make the cardinal/ordinal effect disappear testifies to the inescapable influence of non-mathematical knowledge.

Nevertheless, understanding the determinants of problems' representations is a crucial step in identifying the potential pitfalls and dead ends born from unsuitable representations, as well as helping to develop transfer or learning in a school setting (Gros & Gvozdic, 2022). In this perspective, interventions specifically targeting representational change in arithmetic word problems provide rich insights into the benefits of representation-oriented teaching (Fischer et al., 2019; Gvozdic & Sander, 2020; Iacono et al., 2022). From this angle, the use of drawing production may be a promising path to promote the semantic recoding of a sub-optimal mental representation. Indeed, while insufficient in this experiment to overcome the cardinal-ordinal distinction, drawing tasks could nevertheless constitute one of the potential levers that children can use to help promote cognitive flexibility in the classrooms. Considering the beneficial influence of drawings as a study material – provided that there is adequate instructional support (see Wu & Rau, 2019 for a discussion) – it might be possible to guide students in using drawings to reach a better understanding of the mathematical relations depicted in a problem statement.

Beyond that, prompting children to make drawings using specific features (e.g., representing a cardinal problem using graduated axes), may help target a specific difficulty and foster a new problem representation that would be compatible with a better solving strategy. While prompting drawing production has sometimes yielded mixed results (see Verschaffel, 2016, for a discussion), this line of interventions using teacher-imposed constraints on the drawings has also shown encouraging results (Jitendra & Hoff, 1996; Ng & Lee, 2005; Verschaffel et al., 2020). More work is needed to better understand the role of drawings in promoting representational change using systematic drawing analysis. This line of research may find renewed inspiration in the recent methodological advances leveraging crowd-sourced drawing

analysis to get a deeper understanding of their underlying mental representations (Bainbridge, 2022). Be it through guided or unguided tasks, it is our hope that future research will identify the optimal conditions to make the most of drawing production, in order to fully harness the thousand words' worth of every drawing.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13421-024-01523-w>.

Funding This research was supported by grants from the Regional Council of Burgundy, Paris Feder Grants (20159201AAO050S02982 & 20169201AAO050S01845), the Experimental Fund for the Youth and French Ministry of Education (HAP10-CRE-EXPE-S1), and from the French Ministry of Education and Future Investment Plan (CS-032-15-836-ARITHM-0).

Data availability All data and materials for this study are available online at https://osf.io/nhp4e/?view_only=6a5e4fd4dedb4811a994030b988033a8.

Declarations

Conflicts of interest We have no conflicts of interest to disclose.

Ethics approval This research was approved by the ethics committee of the University of Geneva (decision n° PSE.20181104.18).

Consent to participate All participants gave informed consent to participate in the study.

References

- Agrell, B., & Dehlin, O. (1998). The clock-drawing test. *Age and ageing*, 27(3), 399–404. <https://doi.org/10.1093/ageing/afs149>
- Anderson, S. F., & Kelley, K. (2018). BUCSS: Bias and Uncertainty Corrected Sample Size [computer software and manual]. R package version 1.2.1, <https://CRAN.R-project.org/package=BUCSS>
- Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28(11), 1547–1562
- Baccaglioni-Frank, A., Carotenuto, G., & Sinclair, N. (2020). Eliciting preschoolers' number abilities using open, multi-touch environments. *ZDM*, 52, 779–791.
- Bagnoud, J., Burra, N., Castel, C., Oakhill, J., & Thevenot, C. (2018). Arithmetic word problems describing discrete quantities: EEG evidence for the construction of a situation model. *Acta Psychologica*, 190, 116–121.
- Bainbridge, W. A. (2022). A tutorial on capturing mental representations through drawing and crowd-sourced scoring. *Behavior Research Methods*, 54(2), 663–675.
- Bainbridge, W. A., Hall, E. H., & Baker, C. I. (2019). Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nature communications*, 10(1), 5.
- Bakar, K. A., Way, J., & Bobis, J. (2016). *Young children's drawings in problem solving*. Mathematics Education Research Group of Australasia.
- Barrios, F. M. G., & Martínez, E. C. (2014). Diagrams produced by secondary students in multiplicative comparison word problems. *Journal of Mathematics and System Science*, 4(2), 83.
- Bassok, M. (2001). Semantic alignments in mathematical word problems. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 401–433). MIT Press.
- Berkum, J. J. V., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657–671.
- Bermejo, V. (1996). Cardinality development and counting. *Developmental Psychology*, 32(2), 263–268.
- Burgess, A. W., & Hartman, C. R. (1993). Children's drawings. *Child Abuse & Neglect*, 17(1), 161–168.
- Carruthers, E., & Worthington, M. (2003). *Children's mathematics: making marks, making meaning*. Sage.
- Chamberlain, R., & Wagemans, J. (2016). The genesis of errors in drawing. *Neuroscience & Biobehavioral Reviews*, 65, 195–207.
- Cheung, C. N., & Lourenco, S. F. (2019). Does $1 + 1 = 2$ nd? The relations between children's understanding of ordinal position and their arithmetic performance. *Journal of Experimental Child Psychology*, 187, 104651.
- Colomé, Á., & Noël, M. P. (2012). One first? Acquisition of the cardinal and ordinal uses of numbers in preschoolers. *Journal of Experimental Child Psychology*, 113(2), 233–247.
- Coquin-Viennot, D., & Moreau, S. (2003). Highlighting the role of episodic model in the solving of arithmetical problems. *European Journal of Psychology and Education*, 18, 267–279.
- Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, 9(4), 343–363.
- Crespo, S. M., & Kyriakides, A. O. (2007). Research, Reflection, Practice: To Draw or Not to Draw: Exploring Children's Drawings for Solving Mathematics Problems. *Teaching Children Mathematics*, 14(2), 118–125.
- Csikós, C., Sztányi, J., & Kelemen, R. (2012). The effects of using drawings in developing young children's mathematical word problem solving: A design experiment with third-grade Hungarian students. *Educational Studies in Mathematics*, 81(1), 47–65.
- Cummins, D. D. (1991). Children's interpretations of arithmetic word problems. *Cognition and Instruction*, 8(3), 261–289.
- Dantzig, T. (1945). *Number the language of science. A critical survey written for the cultured non-mathematician*. The Macmillan Company.
- Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, 6, 348.
- Dauben, J. W. (1990). *Georg Cantor: His mathematics and philosophy of the infinite*. Princeton University Press.
- De Bock, D., & Verschaffel, Janssens D. (1998). The predominance of linear model in secondary students' solutions of word problems involving length and area of similar plane figures. *Education Studies in Mathematics*, 35(1), 65–83.
- De Bock, D., Verschaffel, L., Janssens, D., Van Dooren, W., & Claes, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modelling non-linear geometry problems. *Learning and Instruction*, 13(4), 441–463.
- De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solution. *Journal of Educational Psychology*, 77, 460–470.
- Draschkow, D., Wolfe, J. M., & Vo, M. L. H. (2014). Seek and you shall remember: Scene semantics interact with visual search to build better memories. *Journal of Vision*, 14(8), 10–10.
- Edens, K., & Potter, E. (2007). The relationship of drawing and mathematical problem solving: Draw for math tasks. *Studies in Art Education*, 48(3), 282–298.
- Edens, K., & Potter, E. (2008). How students “unpack” the structure of a word problem: Graphic representations and problem solving. *School Science and Mathematics*, 108(5), 184–196.

- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215.
- Fan, J. E. (2015). Drawing to learn: How producing graphical representations enhances scientific thinking. *Translational Issues in Psychological Science*, 1(2), 170.
- Fan, J. E., Hawkins, R. D., Wu, M., & Goodman, N. D. (2020). Pragmatic inference and visual abstraction enable contextual flexibility during visual communication. *Computational Brain & Behavior*, 3, 86–101.
- Fischer, F. E., & Beckey, R. D. (1990). Beginning kindergarteners' perception of number. *Perceptual and Motor Skills*, 70(2), 419–425.
- Fischer, J. P., Sander, E., Sensevy, G., Vilette, B., & Richard, J. F. (2019). Can young students understand the mathematical concept of equality? A whole-year arithmetic teaching experiment in second grade. *European Journal of Psychology of Education*, 34(2), 439–456.
- Fourtassi, M., Rode, G., & Pisella, L. (2017). Using eye movements to explore mental representations of space. *Annals of Physical and Rehabilitation Medicine*, 60(3), 160–163.
- Freeman, N. H., & Janikoun, R. (1972). Intellectual realism in children's drawings of a familiar object with distinctive features. *Child Development*, 1116–1121. <https://doi.org/10.2307/1127668>
- Frege, G. (1984). *The Foundations of Arithmetic* (J. Austin, Trans.) Northwestern University Press.
- Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit non-linguistic task. *Cognitive Science*, 34(8), 1430–1451.
- Gainotti, G., & Trojano, L. (2018). Constructional apraxia. *Handbook of Clinical Neurology*, 151, 331–348.
- Gamo, S., Sander, E., & Richard, J.-F. (2010). Transfer of strategy use by semantic recoding in arithmetic problem solving. *Learning and Instruction*, 20(5), 400–410.
- Geary, D. C. (2018). Growth of symbolic number knowledge accelerates after children understand cardinality. *Cognition*, 177, 69–78.
- Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2018). Early conceptual understanding of cardinality predicts superior school-entry number-system knowledge. *Psychological science*, 29(2), 191–205.
- Goffin, C., & Ansari, D. (2016). Beyond magnitude: Judging ordinality of symbolic number is unrelated to magnitude comparison and independently relates to individual differences in arithmetic. *Cognition*, 150, 68–76.
- Gros, H., & Gvozdic, K. (2022). Context, content effects and flexibility. In E. Clément (Eds.), *Cognitive Flexibility: The Cornerstone of Learning* (pp. 203–232). <https://doi.org/10.1002/9781119902737.ch8>
- Gros, H., Sander, E., & Thibaut, J. P. (2019). When masters of abstraction run into a concrete wall: Experts failing arithmetic word problems. *Psychonomic Bulletin & Review*, <https://doi.org/10.3758/s13423-019-01628-3> online first.
- Gros, H., Sander, E., & Thibaut, J. P. (2020a). Are content effects out of sight? An eye-tracking study of arithmetic problem solving. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. Cognitive Science Society.
- Gros, H., Thibaut, J. P., & Sander, E. (2020b). Semantic congruence in arithmetic: A new conceptual model for word problem solving. *Educational Psychologist*, 55(2), 69–87.
- Gros, H., Thibaut, J. P., & Sander, E. (2021). What we count dictates how we count: A tale of two encodings. *Cognition*, 212, 104665.
- Gvozdic, K., & Sander, E. (2020). Learning to be an opportunistic word problem solver: Going beyond informal solving strategies. *ZDM*, 52(1), 111–123.
- Hebart, M. N., Zheng, C. Y., Pereira, F., & Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behaviour*, 4(11), 1173–1185.
- Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology*, 91(4), 684.
- Hobson, S. M., Trundle, K. C., & Saçkes, M. (2010). Using a planetary software program to promote conceptual change with young children. *Journal of Science Education and Technology*, 19, 165–176.
- Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. *Child Development*, 54(1), 84–90. <https://doi.org/10.2307/112986>
- Hund, A. M., Kannass, K. N., Bove, R., Fairweather, L., Maydew, M., & Monla, A. (2021). Young children's understanding of ordinal and spatial labels. *Cognitive Development*, 58, 101041.
- Iacono, E., Gros, H., & Clément, E. (2022). Training flexible categorization to improve arithmetic problem solving: A school-based intervention with 5th graders. In Culbertson, J., Perfors, A., Rabagliati, H., & Ramenzoni, V. (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*. Cognitive Science Society (Vol. 44).
- Jitendra, A. K., & Hoff, K. (1996). The effects of schema-based instruction on the mathematical word-problem solving performance of students with learning disabilities. *Journal of Learning Disabilities*, 29, 422–432.
- Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, 107(43), 18243–18250. <https://doi.org/10.1073/pnas.1012933107>
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92(1), 109.
- Kosslyn, S. M., Murphy, G. L., Bemesderfer, M. E., & Feinstein, K. J. (1977). Category and continuum in mental comparisons. *Journal of Experimental Psychology: General*, 106(4), 341.
- Lakoff, G., & Núñez, R. E. (2000). Where mathematics comes from: How the embodied mind brings mathematics into being. *AMC*, 10(12), 720–733.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–99.
- Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105(2), 395–438.
- Lenth, R. V., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2023). *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*; R Package Version 1.8.9; 2023.
- Lewis-Peacock, J. A., Drysdale, A. T., & Postle, B. R. (2015). Neural evidence for the flexible control of mental representations. *Cerebral Cortex*, 25(10), 3303–3313.
- Long, B., Fan, J. E., & Frank, M. C. (2018). Drawings as a window into developmental changes in object representations. In Kalish, C., Rau, M., Zhu, J., & Rogers, T. T. (Eds.), *Proceedings of the 40th Annual Conference of Cognitive Science Society* (pp. 708–713). Madison, WI: Cognitive Science Society.
- Makuuchi, M., Kaminaga, T., & Sugishita, M. (2003). Both parietal lobes are involved in drawing: a functional MRI study and implications for constructional apraxia. *Cognitive Brain Research*, 16(3), 338–347.
- Martin, S. A., & Bassok, M. (2005). Effects of semantic cues on mathematical modeling: Evidence from word-problem solving and equation construction tasks. *Memory & Cognition*, 33, 471–478.
- Meade, M. E., Wammes, J. D., & Fernandes, M. A. (2018). Drawing as an encoding tool: Memorial benefits in younger and older adults. *Experimental Aging Research*, 44(5), 369–396.

- Meyer, C., Barbiere, S., & Weerman, F. (2016). Order and Ordinality: The Acquisition of Cardinals and Ordinals in Dutch. *Boston University Conference on Language Development (BUCLD)* (40th ed.). Cascadia.
- Mikkilä-Erdmann, M., Södervik, I., Vilppu, H., Käpä, P., & Olkinuora, E. (2012). First-year medical students' conceptual understanding of and resistance to conceptual change concerning the central cardiovascular system. *Instructional Science*, 40, 745–754.
- Miller, S. E., Marcovitch, S., Boseovski, J. J., & Lewkowicz, D. J. (2015). Young children's ability to use ordinal labels in a spatial search task. *Merrill-Palmer Quarterly*, 61(3), 345–361.
- Miller, K., Major, S. M., Shu, H., & Zhang, H. (2000). Ordinal knowledge: Number names and number concepts in Chinese and English. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 54(2), 129.
- Ng, S.-F., & Lee, K. (2005). How primary five pupils use the model method to solve word problems. *The Mathematics Educator*, 9(1), 60–83.
- Orrantia, J., & Muñoz, D. (2013). Arithmetic word problem solving: evidence for a magnitude-based mental representation. *Memory & Cognition*, 41, 98–108.
- Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences*, 112(33), 10089–10092.
- Pinnegar, S., Mangelson, J., Reed, M., & Groves, S. (2011). Exploring preservice teachers' metaphor plotlines. *Teaching and Teacher Education*, 27(3), 639–647.
- Reeve, R. A. (1996). The Referential Adequacy of Students' Visual Analogies of Fractions. *Mathematical Cognition*, 2(2), 137–169. <https://doi.org/10.1080/135467996387507>
- Rellensmann, J., Schukajlow, S., & Leopold, C. (2017). Make a drawing. Effects of strategic knowledge, drawing accuracy, and type of drawing on students' mathematical modelling performance. *Educational Studies in Mathematics*, 95, 53–78.
- Reusser, K. (1990). From text to situation to equation: Cognitive simulation of understanding and solving mathematical word problems. In H. Mandl, E. De Corte, N. Bennet, & H. F. Friedrich (Eds.), *Learning and Instruction, European Research in an International Context* (II). Pergamon Press.
- Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. *The Development of Mathematical Thinking* (pp. 153–196). New York: Academic Press.
- Roberts, B. R., & Wammes, J. D. (2021). Drawing and memory: Using visual production to alleviate concreteness effects. *Psychonomic Bulletin & Review*, 28, 259–267.
- Russell, B. (1919). *Introduction to mathematical philosophy*. Allen & Unwin.
- Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology*, 103(3), 325–337.
- Scheibling-Sève, C., Pasquinelli, E., & Sander, E. (2020). Assessing conceptual knowledge through solving arithmetic word problems. *Educational Studies in Mathematics*, 103(3), 293–311.
- Scheibling-Sève, C., Gvozdic, K., Pasquinelli, E., & Sander, E. (2022). Enhancing cognitive flexibility through a training based on multiple categorization: Developing proportional reasoning in primary school. *Journal of Numerical Cognition*, 8(3), 443–472.
- Schmeck, A., Mayer, R. E., Opfermann, M., Pfeiffer, V., & Leutner, D. (2014). Drawing pictures during learning from scientific text: Testing the generative drawing effect and the prognostic drawing effect. *Contemporary Educational Psychology*, 39(4), 275–286.
- Sharp, E., & Shih Dennis, M. (2017). Model drawing strategy for fraction word problem solving of fourth-grade students with learning disabilities. *Remedial and Special Education*, 38(3), 181–192.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703.
- Shulman, K. I. (2000). Clock-drawing: is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry*, 15(6), 548–561.
- Shusterman, A., Slusser, E., Halberda, J., & Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLOS ONE*, 11(4), e0153072. <https://doi.org/10.1371/journal.pone.0153072>
- Silver, R. (2009). Identifying children and adolescents with depression: Review of the stimulus drawing task and draw a story research. *Art Therapy*, 26(4), 174–180.
- Suppes, P. (1972). *Axiomatic set theory*. Dover Publications.
- Thevenot, C. (2010). Arithmetic word problem solving: Evidence for the construction of a mental model. *Acta Psychologica*, 133(1), 90–95.
- Thevenot, C., & Barrouillet, P. (2015). Arithmetic word problem solving and mental representations. In R. C. Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 158–179).
- Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representations in arithmetic word problem solving. *The Quarterly Journal of Experimental Psychology Section A*, 58(7), 1311–1323.
- Thevenot, C., & Oakhill, J. (2006). Representations and strategies for solving dynamic and static arithmetic word problems: The role of working memory capacities. *European Journal of Cognitive Psychology*, 18(5), 756–775.
- Thom, J. S., & McGarvey, L. M. (2015). The act and artifact of drawing (s): observing geometric thinking with, in, and through children's drawings. *ZDM*, 47, 465–481.
- Thomas, G. V., & Jolley, R. P. (1998). Drawing conclusions: A re-examination of empirical and conceptual bases for psychological evaluation of children from their drawings. *British Journal of Clinical Psychology*, 37(2), 127–139.
- Trundle, K., Atwood, R., & Christopher, J. (2007). Fourth-grade elementary students' conceptions of standards-based lunar concepts. *International Journal of Science Education*, 29(5), 595–616.
- Tunç-Pekkan, Z. (2015). An analysis of elementary school children's fractional knowledge depicted with circle, rectangle, and number line representations. *Educational Studies in Mathematics*, 89, 419–441.
- Ucar, S., Trundle, K. C., & Krissek, L. (2011). Inquiry-based instruction with archived, online data: An intervention study with preservice teachers. *Research in Science Education*, 41, 261–282.
- Uesaka, Y., Manalo, E., & Ichikawa, S. I. (2010). The effects of perception of efficacy and diagram construction skills on students' spontaneous use of diagrams when solving math word problems. *Diagrams 2010, Portland, OR, USA, August 9-11, 2010* (pp. 197–211). Springer Berlin Heidelberg.
- Van Essen, G. V., & Hamaker, C. (1990). Using self-generated drawings to solve arithmetic word problems. *The Journal of Educational Research*, 83(6), 301–312.
- Van Garderen, D., & Montague, M. (2003). Visual-spatial representation, mathematical problem solving, and students of varying abilities. *Learning Disabilities Research & Practice*, 18(4), 246–254.
- Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational psychology review*, 17, 285–325.
- Verschaffel, L. (2016). Get the picture? On the role of graphical representations in the solution of mathematical word problems. In *Plenary lecture at the Conference of the EARLI SIG 2, Comprehension of Text and Graphics, University of Geneva, Geneva, Switzerland*.
- Verschaffel, L., De Corte, E., & Vierstraete, H. (1999). Upper elementary school pupils' difficulties in modeling and solving nonstandard additive word problems involving ordinal numbers. *Journal for Research in Mathematics Education*, 30(3), 265–285.

- Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM*, *52*, 1–16.
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*(4), 535–585.
- Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *Quarterly Journal of Experimental Psychology*, *69*(9), 1752–1776.
- Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2017). Learning terms and definitions: Drawing and the role of elaborative encoding. *Acta Psychologica*, *179*, 104–113.
- Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2018). Creating a recollection-based memory through drawing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(5), 734.
- Wasner, M., Moeller, K., Fischer, M. H., & Nuerk, H. C. (2015). Related but not the same: Ordinality, cardinality and 1-to-1 correspondence in finger-based numerical representations. *Journal of Cognitive Psychology*, *27*(4), 426–441.
- Wechsler, D. (2009). *WMS-IV: Wechsler memory scale* (4th ed.). Pearson.
- Weinman, J., Petrie, K. J., Moss-Morris, R., & Horne, R. (1996). The illness perception questionnaire: a new method for assessing the cognitive representation of illness. *Psychology and Health*, *11*(3), 431–445.
- Westenskow, A., Moyer-Packenham, P., Anderson-Pence, K., Shumway, J., & Jordan, K. (2014). Cute drawings? The disconnect between students' pictorial representations and their mathematics responses to fraction questions. *Revista Internacional de Pesquisa em Educação Matemática*, *4*(1), 81–105.
- Wu, S. P., & Rau, M. A. (2019). How students learn content in science, technology, engineering, and mathematics (STEM) through drawing activities. *Educational Psychology Review*, *31*, 87–120.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750.
- Yoshida, H., & Shinmachi, Y. (1999). The influence of instructional intervention on children's understanding of fractions. *Japanese Psychological Research*, *41*(4), 218–228.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Practices Statement All data and materials for this study are available online on OSF at https://osf.io/nhp4e/?view_only=6a5e4fd4dedb4811a994030b988033a8.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.