



HAL
open science

What we count dictates how we count : a tale of two encodings

Hippolyte Gros

► **To cite this version:**

Hippolyte Gros. What we count dictates how we count : a tale of two encodings. Cognitive Sciences. Université Paris Cité, 2019. English. NNT : 2019UNIP5016 . tel-03554724v2

HAL Id: tel-03554724

<https://theses.hal.science/tel-03554724v2>

Submitted on 3 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université de Paris

FIRE Doctoral School – Frontières de l’Innovation en Recherche et Éducation

What we count dictates how we count
A tale of two encodings

THÈSE

Présentée par Hippolyte Gros

En vue de l’obtention du Doctorat en Sciences Cognitives
Label Doctorat Européen (Doctor Europaeus)

Sous la direction de

Emmanuel Sander – Professeur à L’Université de Genève

Et la codirection de

Jean-Pierre Thibaut – Professeur à l’Université de Bourgogne Franche-Comté

Présentée et soutenue publiquement le 15 novembre 2019

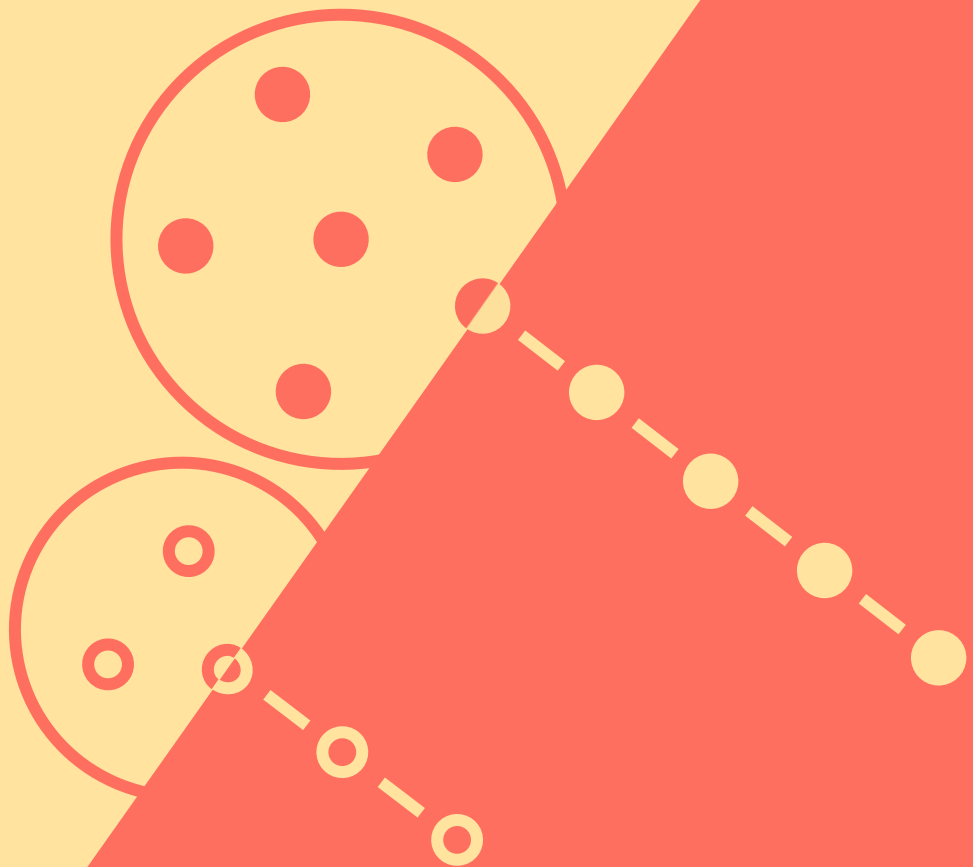
Devant un jury composé de :

Catherine Thevenot	Université de Lausanne	Rapportrice
Lieven Verschaffel	KU Leuven	Rapporteur
Évelyne Clément	Université Cergy-Pontoise	Examinatrice
Véronique Izard	Université Paris Descartes	Examinatrice
Emmanuel Sander	Université de Genève	Directeur
Jean-Pierre Thibaut	Université Bourgogne Franche-Comté	Co-directeur



Except where otherwise noted, this is work licensed under
<https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>

WHAT WE COUNT
DICTATES
HOW WE COUNT



2
A
TALE OF
ENCODINGS

ABSTRACT

With its context-independent rules valid in any setting, mathematics is considered to be the champion of abstraction, and for a long time human mathematical reasoning was thought to follow nothing but the laws of logic. However, the idea that mathematics is grounded in nature has gained traction over the past decades, and the context-independency of mathematical reasoning has come to be questioned. The thesis we defend concerns the role played by general, non-mathematical knowledge on individuals' understanding of numerical situations. We propose that *what* we count has a crucial impact on *how* we count, in the sense that human's representation of numerical information is dependent on the semantic context in which it is embedded. More specifically, we argue that general, non-mathematical knowledge about the entities described in a mathematical word problem can shape its interpretation and foster one of two representations: either a cardinal encoding, or an ordinal encoding.

After introducing a new framework of arithmetic word problem solving accounting for the interactions between mathematical knowledge and world knowledge in the encoding, recoding and solving of arithmetic word problems, we present a series of 16 experiments assessing how world knowledge about specific quantities can promote one of two problem representations. Using isomorphic arithmetic word problems involving either cardinal quantities (weights, prices, collections) or ordinal quantities (durations, heights, number of floors), we investigate the pervasiveness of the cardinal-ordinal distinction in a wide range of activities, including problem categorization, problem comparison, algorithm selection, problem solvability assessment, problem recall, sentence recognition, drawing production and transfer of strategies. We gather data using behavioral measures (success rates, algorithm use, response times) as well as eye tracking (fixation times, saccades, pupil dilation), to show that the difference between problems meant to foster either a cardinal or an ordinal encoding has a far-reaching influence on participants from diverse populations ($N = 2180$), ranging from 2nd graders and 5th graders to lay adults, expert mathematicians and math teachers.

We discuss the general educational implications of these effects of semantic (in)congruence, and we propose new directions for future research on this crucial issue. We conclude that these findings illustrate the extent to which human reasoning is constrained by the content on which it operates, even in domains where abstraction is praised and trained.

RÉSUMÉ

Parce qu'elles manipulent des objets fondamentalement abstraits, les lois mathématiques ont une validité indépendante du contexte dans lequel elles s'appliquent. Autrement dit, $2 + 2 = 4$, que l'on compte des pommes, des schtroumpfs, ou des années-lumière. Par extension, il a longtemps été considéré qu'il en était de même pour la pensée mathématique chez l'humain, perçue comme objective et indépendante des contenus sur lesquels elle s'exerce. Pourtant, un nombre grandissant de travaux s'accordent à dire que la logicité n'est pas seule à gouverner la pensée humaine, que le contexte dans lequel il se trouve influence ses raisonnements, et que la pensée mathématique est fondamentalement incarnée. Ainsi, notre thèse est que les connaissances générales des individus influencent considérablement leurs représentations des situations numériques. En particulier, nous faisons l'hypothèse que les savoirs non-mathématiques des individus au sujet des entités décrites dans un problème peuvent façonner leur représentation de la situation, les poussant à en réaliser un encodage soit cardinal, soit ordinal.

Nous commençons par présenter un modèle conceptuel visant à décrire les interactions entre la sémantique du monde et la sémantique mathématique évoquées à la lecture d'un problème arithmétique à énoncé verbal. Nous faisons la prédiction que les connaissances générales sur le monde influent sur l'encodage, le recodage et la résolution des problèmes arithmétiques à énoncés verbaux, notamment en induisant des représentations soit cardinales, soit ordinales. Nous évaluons cette hypothèse grâce à 16 expériences fondées sur l'étude d'énoncés isomorphes implémentés avec certaines entités censées susciter un encodage cardinal (poids, prix, collections d'éléments) ou ordinal (durées, hauteurs, nombre d'étages). Nous montrons la robustesse de ces effets au travers d'une variété de tâches, qu'il s'agisse de classification, comparaison, résolution, production graphique, jugement de solubilité, évaluation de solution, reconnaissance, transfert et rappel de problèmes. La prévalence des effets observés est déterminée par des indices comportementaux (performances, temps de réponse, sélection de stratégies) et physiologiques (oculométrie et pupillométrie), collectés auprès d'enfants du CE1 au CM2, ainsi que d'adultes tout venants, d'enseignants en mathématiques et d'experts mathématiciens. Les riches enjeux éducatifs portés par ces questions sont discutés de même que les perspectives ouvertes par la prise en compte des effets de congruence sémantique. Nous concluons sur les contraintes que les contenus opèrent sur le raisonnement.

REMERCIEMENTS

À l’instar du raisonnement humain qui, comme nous le verrons au cours de cette thèse, est influencé par le cadre dans lequel il opère, tout travail de thèse est façonné par le contexte dans lequel il est réalisé. La forme finale de ce manuscrit est donc aussi, en partie, le produit des différentes personnes qui m’ont apporté leur expertise, leurs conseils et leur soutien lors de ces quatre dernières années. Je tiens à les en remercier chaleureusement :

Mes premiers remerciements vont à Emmanuel Sander, qui m’a apporté sa confiance dès nos premiers échanges, en me proposant un projet de recherche qui fut le « pilote » de cette thèse. Son exemple, en termes d’érudition, de créativité, d’exigence intellectuelle et de vivacité d’esprit demeure une source constante d’inspiration. Je le remercie pour les nombreuses opportunités qu’il a su créer pour moi, à Paris puis à Genève, mais aussi pour la patience, la hauteur de vue et la bienveillance avec lesquelles il a toujours encadré mon travail. Sur le plan scientifique comme sur le plan humain, je tiens enfin à souligner le réel plaisir que m’ont procuré ces années de collaboration. Je crois avoir beaucoup appris à son contact, et je me réjouis continuellement d’avoir la chance de pouvoir travailler et apprendre à ses côtés.

Je remercie également très chaleureusement Jean-Pierre Thibaut, pour son co-encadrement toujours disponible, même à distance, pour la rigueur de ses méthodes, la justesse de ses retours et la grande sagacité de ses intuitions, mais aussi pour sa capacité à pointer du doigt les limites de mes raisonnements, en posant les questions difficiles qui amènent à réellement progresser. Il va sans dire que cette thèse n’aurait pas été la même si je n’avais pu bénéficier de son expérience, et je lui en suis grandement reconnaissant.

J’adresse aussi mes remerciements à Catherine Thévenot et Lieven Verschaffel, pour l’honneur qu’ils m’ont fait en acceptant d’être les rapporteur-riche-s de cette thèse. J’ai éprouvé un franc plaisir à lire et à relire leurs travaux respectifs au cours de ces derniers années, et je me réjouis sincèrement de pouvoir bénéficier de leurs commentaires éclairés sur mon travail.

Un grand merci également à Évelyne Clément et Véronique Izard, qui ont accepté d’être examinatrices de cette thèse, et qui m’ont toutes deux prodigué un précieux soutien au cours de ces quatre années. Évelyne, par ses suggestions bienveillantes, averties et pénétrantes au cours de nos nombreux *team meetings*, Véronique, par les retours rigoureux et perspicaces qu’elle m’a offerts en tant que membre de mon *Thesis Advisory Committee*. Merci aussi à Franck Ramus, également membre du *Thesis Advisory Committee*, pour la puissance critique de ses observations et la confiance qu’il m’a toujours inspirée envers la démarche scientifique.

Mes remerciements s’adressent également à Katarina Gvozdić, « co-doctorante » exemplaire et acolyte infatigable, pour ses nombreuses corrections linguistiques, ses relectures attentives à toute heure du jour et de la nuit, ses conseils ingénieux, son énergie contagieuse, son soutien moral, sa force de proposition, sa vive intelligence ainsi que sa générosité naturelle. Merci aussi à elle pour nos aventures européennes, passées et à venir.

Merci bien sûr aux membres de notre séminaire de (post-)doctorants : Lucas Raynal, Sébastien Puma, Emmanuel Trouche, Calliste Scheibling-Sève et David Piot, pour leurs nombreux retours, pour la convivialité de nos échanges et le caractère toujours stimulant de nos interactions. Les discussions scientifiques sur nos objets de recherche respectifs furent l'un des aspects les plus gratifiants de mon expérience de thèse, et je les en remercie tou-te-s chaleureusement.

Merci également aux collègues de notre équipe genevoise, Géry Marcoux, Katia Lehraus, Walther Tessaro et Alexia Forget, pour la sincère cordialité de leur accueil et la richesse des collaborations qui se dessinent.

Je souhaite aussi exprimer ma gratitude à l'école doctorale des Frontières de l'Innovation en Recherche et Éducation, pour avoir financé les deux premières années de cette thèse, et pour avoir créé les opportunités permettant à une communauté de doctorant-e-s interdisciplinaires de se rencontrer. Merci donc aux esprits affûtés que j'ai eu le plaisir d'y côtoyer. Pour les conseils en analyses statistiques, en réalisations graphiques, en sélection de couleurs, ainsi que pour les riches perspectives interdisciplinaires explorées ensemble, un grand merci à Maxime Maheu, J Boussier, Ignacio Rebollo, Jo Zelefsky, Camilla Scaff, Clara Dussaux, Jennifer Tussaut et Elsa Cavatas.

Merci également à la constellation de jeunes et moins jeunes chercheurs en sciences cognitives rencontrée au Cogmaster, pour la richesse de nos échanges, la vivacité de nos débats, et pour avoir probablement contribué à ma culture scientifique à part égale avec notre formation officielle. Merci en particulier à Ralitsa Todorova, Arnaud Beauny, Vladimir Albert, Mehdi Boudjadja, Lucie Berkovitch, Justine Mertz, Anna Kolesnikova, Paloma Jeretic, Timothée Behra, Guillaume Legendre, Clara Arnoux, Anouk Dieuleveut, et Guillaume de Longuemar. C'est notamment en rencontrant cette communauté d'individus passionnés que j'ai confirmé mon intérêt pour le monde de la recherche, et je suis aussi reconnaissant aux responsables de cette formation d'avoir su créer un environnement si propice à ces échanges.

Merci également aux membres du laboratoire Paragraphe qui ont accompagné mes deux premières années de thèse à Paris. Merci à Marianne Habib, Sabine Guéraud, Louise Goyet, Aurélie Coubart et Amélie Lesenecal, d'avoir été des collègues si chaleureuses.

Sur une note plus personnelle, merci à Antoine Gaudin, pour m'avoir précédé dans l'expérience de la thèse et m'avoir montré qu'il était possible de finir un doctorat sans trop sacrifier de sa santé mentale. Merci, entre autres, pour le soutien logistique et moral, mais merci, surtout, pour 10 belles années d'aventures.

Merci à Milo Flouw pour ses suggestions graphiques inspirées, son soutien moral et son accueil toujours chaleureux. Merci à Coline Gros et Benjamin Flouw pour les mêmes raisons.

Merci, enfin, à mes parents, pour leur soutien inconditionnel, et pour toutes les choses qu'ils m'ont apportées. Il faudrait un autre manuscrit de thèse, en plus de celui-ci, pour toutes les énumérer.

It was a fun ride.

TABLE OF CONTENTS

ABSTRACT.....	3
RÉSUMÉ.....	4
REMERCIEMENTS	5
TABLE OF CONTENTS.....	7
RELATED SCIENTIFIC OUTPUT.....	13
GENERAL INTRODUCTION	15
PREAMBLE	16
ILLOGICAL CHILDREN AND RATIONAL ADULTS?.....	16
CONTENT SHAPING THOUGHTS AND PROBLEM SOLVING	20
CURRENT UNDERTAKING	23
REFERENCES.....	28
CHAPTER 1 – SEMANTIC CONGRUENCE IN ARITHMETIC: A NEW CONCEPTUAL MODEL FOR WORD PROBLEM SOLVING	35
PRESENTATION.....	36
ABSTRACT.....	37
KEYWORDS.....	37
INTRODUCTION.....	38
ARITHMETIC WORD PROBLEM SOLVING THEORIES	39
THE SCHEMA MODEL.....	39
THE SITUATION MODEL APPROACH	41
THE SEMANTIC ALIGNMENT CONTRIBUTION	42
ALTERNATIVE ENCODINGS AND RE-REPRESENTATION	44
THE SEMANTIC CONGRUENCE (SECO) MODEL.....	45
COMPONENTS	46
PROCESSES.....	48
INNER WORKINGS	49
ACCOUNTING FOR EXISTING RESULTS: CASE STUDIES.....	50
WORLD SEMANTICS ISSUES	51
RECODING ISSUES.....	58
REWORDING ISSUES.....	65
CONCLUSION	71
SECO'S ADDED VALUE	71
SEMANTIC CONGRUENCE AS AN EDUCATIONAL LEVER TO TACKLE ARDUOUS NOTIONS	74
GAINING EXPERTISE	76

BROADER APPLICATION OF THE SECO MODEL	78
CONCLUSION	81
REFERENCES.....	82
CHAPTER 2 – <i>WHAT WE COUNT DICTATES HOW WE COUNT: A TALE OF TWO ENCODINGS</i>.....	93
PRESENTATION.....	94
ABSTRACT.....	96
KEYWORDS.....	96
INTRODUCTION.....	97
AN ONTOLOGICAL DISTINCTION BETWEEN CARDINAL AND ORDINAL SITUATIONS	97
INTERPRETATIVE EFFECTS AT PLAY IN ARITHMETIC PROBLEM SOLVING	99
SEMANTIC INDUCTION IN QUESTION	101
CARDINAL AND ORDINAL REPRESENTATIONS IN ARITHMETIC PROBLEMS	103
CURRENT STUDY	105
OVERVIEW OF THE EXPERIMENTS.....	105
OVERVIEW OF THE SELECTED ORDINAL QUANTITIES	107
OVERVIEW OF THE SELECTED CARDINAL QUANTITIES.....	108
EXPERIMENT 1	109
METHOD.....	110
RESULTS	113
DISCUSSION.....	114
EXPERIMENT 2	115
METHODS	115
RESULTS	116
DISCUSSION.....	118
EXPERIMENT 3	119
METHODS	119
RESULTS	120
DISCUSSION.....	121
EXPERIMENT 4	121
METHODS	122
RESULTS	124
DISCUSSION.....	125
EXPERIMENT 5	126
METHODS	127
RESULTS	130
DISCUSSION.....	131
EXPERIMENT 6	131
METHODS	132

RESULTS	133
DISCUSSION	133
GENERAL DISCUSSION.....	134
REFERENCES.....	138
CHAPTER 3 - TROUBLE DOWN MEMORY LANE: MISREMEMBRANCE REVEALS ENCODING DIFFERENCES IN ARITHMETIC WORD PROBLEMS.....	145
PRESENTATION.....	146
ABSTRACT.....	147
KEYWORDS	147
INTRODUCTION.....	148
EXPERIMENT 1	155
METHOD	156
RESULTS	157
DISCUSSION	160
EXPERIMENT 2	161
METHOD	162
RESULTS	163
DISCUSSION	165
EXPERIMENT 3	166
METHOD	167
RESULTS	173
DISCUSSION	175
GENERAL DISCUSSION.....	176
REFERENCES.....	179
CHAPTER 4 – WHEN MASTERS OF ABSTRACTION RUN INTO A CONCRETE WALL: EXPERTS FAILING ARITHMETIC WORD PROBLEMS	185
PRESENTATION.....	186
ABSTRACT.....	187
KEYWORDS.....	187
INTRODUCTION.....	188
STUDY 1	190
METHODS.....	190
RESULTS	195
DISCUSSION	197
STUDY 2.....	198
METHODS.....	198
RESULTS	198
DISCUSSION	200

GENERAL DISCUSSION.....	200
ACKNOWLEDGMENTS	202
REFERENCES.....	203
CHAPTER 5 – ARE CONTENT EFFECTS OUT OF TEACHERS’ SIGHT? AN EYE-TRACKING STUDY OF ARITHMETIC WORD PROBLEM SOLVING	207
PRESENTATION.....	208
ABSTRACT.....	209
KEYWORDS.....	209
INTRODUCTION.....	210
SEMANTIC DETERMINANTS OF ARITHMETIC WORD PROBLEM SOLVING	210
CARDINAL VERSUS ORDINAL ENCODING: INVESTIGATING THE CONSTRUCTED REPRESENTATIONS.....	211
EYE TRACKING AS AN INDEX OF REASONING PROCESSES	215
CURRENT STUDY	217
EXPERIMENT 1	218
METHODS	219
RESULTS	220
DISCUSSION.....	223
EXPERIMENT 2	224
METHODS	225
RESULTS	227
DISCUSSION.....	234
GENERAL DISCUSSION.....	237
REFERENCES.....	239
CHAPTER 6 – CAN CHILDREN THINK ‘STRAIGHT’? ORDINALITY AND CARDINALITY IN ARITHMETIC WORD PROBLEM SOLVING.....	247
PRESENTATION.....	248
ABSTRACT.....	249
KEYWORDS.....	249
INTRODUCTION.....	250
IMPORTANCE OF CONTENT EFFECTS THROUGHOUT DEVELOPMENT AND EDUCATION.....	250
INVESTIGATION OF MENTAL REPRESENTATIONS	251
DEVELOPMENT OF CARDINALITY AND ORDINALITY.....	253
CARDINALITY AND ORDINALITY IN ARITHMETIC PROBLEM SOLVING.....	254
PRESENT STUDY	256
EXPERIMENT 1	257
METHODS	257
RESULTS	262
DISCUSSION.....	265

EXPERIMENT 2	266
METHODS.....	267
RESULTS	269
DISCUSSION.....	270
GENERAL DISCUSSION.....	272
REFERENCES.....	274
CHAPTER 7 – ROBUSTNESS OF SEMANTIC ENCODING EFFECTS IN A TRANSFER TASK FOR MULTIPLE-STRATEGY ARITHMETIC PROBLEMS.....	278
PRESENTATION.....	279
ABSTRACT.....	280
KEYWORDS.....	280
INTRODUCTION.....	281
METHODS.....	286
RESULTS	287
DISCUSSION.....	291
REFERENCES.....	293
GENERAL DISCUSSION.....	295
OUR CONTRIBUTION	296
KEY FINDINGS.....	296
ISSUES REQUIRING FURTHER INVESTIGATION	297
THE PARADOX OF P-ITEMS?	299
BEYOND THE DICHOTOMY	302
TAKING A TURN FOR THE WHORF.....	303
HOW CAN TRANSFER BE IMPROVED?.....	305
HYBRID PROBLEMS AS PIVOTAL SITUATIONS?.....	306
ARGUMENTATION TO PROMOTE SEMANTIC RECODING?	308
CONCLUSION	312
REFERENCES.....	313

RELATED SCIENTIFIC OUTPUT

This thesis comprises the following articles that have been accepted, submitted, or are currently in preparation for publication:

Chapter 1.

Gros, H., Thibaut, J.-P., & Sander, E. (*accepted*). Semantic congruence in arithmetic: A new conceptual model for word problem solving. *Educational Psychologist*.

Chapter 2 – General.

Gros, H., Thibaut, J.-P., & Sander, E. (*under review – Journal of Experimental Psychology: General*). What we count dictates how we count: A tale of two encodings.

Chapter 2 – Experiments 5 & 6 only.

Gros, H., Sander, E., & Thibaut, J. P. (2016). “This problem has no solution”: When closing one of two doors results in failure to access any. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1271-1276). Austin, TX: Cognitive Science Society. DOI: 10.13140/RG.2.1.1544.5369

Chapter 3.

Gros, H., Thibaut, J.-P., & Sander, E. (*in prep*). Trouble down memory lane: Misremembrance reveals encoding differences in arithmetic word problems.

Chapter 4.

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*, online first.

Chapter 5.

Gros, H., Sander, E., & Thibaut, J.-P. (*in prep*). Are content effects out of sight for teachers too? An eye-tracking study of arithmetic word problem solving.

Chapter 6 – General.

Gros, H., Thibaut, J.-P., & Sander, E. (*in prep*). Learning how to think “straight”: Development of ordinal representations of arithmetic problems.

Chapter 6 – Experiment 1 only.

Gros, H., Thibaut, J.-P., & Sander, E. (2017). The nature of quantities influences the representation of arithmetic problems: evidence from drawings and solving

procedures in children and adults. In R. Granger, U. Hahn, & R. Sutton (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp 439-444). Austin, TX: Cognitive Science Society.

Chapter 7.

Gros, H., Thibaut, J.-P., & Sander, E. (2015). Robustness of semantic encoding effects in a transfer task for multiple-strategy arithmetic word problems. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 818-823). Austin, TX: Cognitive Science Society. DOI: 10.13140/RG.2.1.5178.4165.

The Introduction and General Discussion of this thesis build on these articles.

GENERAL INTRODUCTION



Preamble

Consider the following argument:

All mammals need water

All PhD students need water

Therefore, all PhD students are mammals

This categorical syllogism is invalid. Even though the conclusion “all PhD students are mammals” is objectively true, it cannot be deduced from the two premises. However, when asked to evaluate syllogisms of the same form, with a true but logically invalid conclusion, participants tend to accept them 2 out of 3 times (Stanovich, 1999). Now consider this new syllogism:

All sea cucumbers need oxygen

All PhD student need oxygen

Therefore, All PhD students are sea cucumbers

Hopefully no one will make the mistake of considering this argument valid. In fact, research has shown that almost everyone agrees that such syllogisms, with a false conclusion, are invalid (Stanovich, 1999). But what accounts for this major difference in performance? These two syllogisms have the same form, their logical structure is the same, and a rational mind should be able to evaluate their validity regardless of whether they mention mammals or sea cucumbers. However, one reaches a conclusion which is true according to our general knowledge, whereas the other reaches a conclusion which is in direct contradiction with what we know about the world. And it appears that our evaluation of abstract ideas is not as removed from concrete life as once believed.

Illogical children and rational adults?

The idea that human reasoning is not exclusively guided by logic is a relatively new development in the history of thought. In the *Nicomachean Ethics*, Aristotle affirms that what distinguishes human beings from animals is their rational principle, or their ability to engage in logically valid reasoning (Aristotle, trans. 1999). For Descartes (trans. 1984, II), God’s gift to man was a reliable intellect, which – if used correctly – was sure to avoid errors in reasoning. Ultimately, for Inhelder and Piaget (1958, p.1), “reasoning is nothing more than propositional

calculus itself". Thus, for an extended period of time, the ability to reason according to the rules of logic was deemed the specificity of the human mind.

Only in the second part of the 20th century has human reasoning started to be the subject of experimental investigation. Until then, it was considered that the study of adult reasoning fell within the scope of logic, hinging on mathematics and philosophy, but not on psychology. In fact, studying human reasoning without leaning on the study of logic formerly seemed like an odd idea, since the two were considered as equivalent. It is no coincidence that Boole gave the title "An investigation of the Laws of Thought" (1854) to his book introducing the rules of Boolean algebra, a formalism for describing logical relations.

Paradoxically, while for many years the logicity of adult reasoning was not called into question and thus not investigated, the opposite was also true for children's reasoning. Aristotle considered children to be profoundly irrational beings, governed by emotions and passions, closer to animals than to a mature man with regards to their rationality (Chamblis, 1982). Even developmental psychology held this view until the late 1970s, under the influence of Piaget's work. In his theory of cognitive development, Piaget considered that the ultimate stage of development, the Formal Operational Stage, was reached around adolescence. At this stage, people were believed to develop the ability to comprehend abstract concepts, and to think logically (Flavell, 1963). In other words, adults were supposed to be experts of logical thinking, give or take the occasional mistake, whereas children were thought to only start engaging in logical and abstract reasoning during the Concrete Operational Stage, between 7 and 11 years old. Thus, the development of logical reasoning from an irrational infant to a fully rational adult is a central element of the Piagetian perspective.

In the last 50 years, significant progress has been made on these questions thanks to experimental psychology, and we have gained a finer understanding of reasoning and abstract thought. Notably, empirical works suggest that (1) human adult reasoning is far from being exclusively guided by logic or mathematical rules, and that (2) children are capable of logic and mathematical reasoning, at least to a certain extent.

Regarding the first point, the work of Wason (1960, 1968) initiated a paradigm shift that opened an entirely new line of research on adults' reasoning biases. He investigated the understanding of conditional statements of the form

“if P then Q ”, by using the now famous *Wason selection task*. In this task, participants are typically presented with four cards, each of which has a number on one side and a letter on the other side. The visible faces of the four cards show, respectively, the numbers “3”, “8”, “A”, and “D”. The following question is presented to the participants “which card(s) must you turn over in order to test the truth of the proposition that if a card shows an even number on one face, then its opposite face has a vowel on it?”. The correct answer to this task is to turn over the “8” card (by *modus ponens*, since 8 is an even number) and the “D” card (by *modus tollens*, since D is not a vowel). However, Wason showed that the correct answer was found by less than 10% of the participants (Wason, 1968). Instead, a majority of adults said they would turn over the “A” card and the “8” card”. This systematic error showed that participants tended to ignore the modus tollens rule (if it is true that P implies Q , and if the contradictory of Q is true, then the contradictory of P is also true). This apparent failure to apply a fundamental rule of propositional logic has been extensively replicated since then, sounding the death knell for the idea that adult reasoning strictly follows the laws of logic (Ragni, Kola, & Johnson-Laird, 2018).

Interestingly, subsequent works showed that participants’ ability to select the correct cards was directly dependent on the situation’s context. For example, when the rather abstract “if even number, then vowel” rule is replaced by a more concrete one, with relevance in a real-life social context, performance increases drastically. Notably, Cosmides and Tooby (1992) showed that using the rule “if you are drinking alcohol, then you must be over 18”, accompanied by the presentations of four cards with, respectively, the inscriptions “16”, “25”, “drinking beer”, and “drinking coke”, most participants have no difficulty in selecting the correct “16” and “drinking beer” cards. In other words, despite adults struggling with the understanding of specific logic relations, their difficulties can be lifted by changing the semantic content of the situation. Several variations of this task have been proposed, and while there is no current consensus regarding the proper explanation for the effects of context they highlight, these effects have been extensively demonstrated and their existence is indisputable (e.g. Cheng & Holyoak, 1985; Cosmides, 1989; Girotto, Kimmelman, Sperber, & Van der Henst, 2001; Cox & Griggs, 1982; Hilton, Kimmelman, & Bonnefon, 2005; Johnson-Laird, Legrenzi, & Legrenzi, 1972; Klauer, Stahl, & Erdfelder, 2007; Manktelow &

Over, 1991; Politzer & Nguyen-Xuan, 1992; Stenning & van Lambalgen, 2001; see Ragni, Kola, & Johnson-Laird, 2018, for a meta-analysis).

Now, regarding the second point we raised – the question of whether children can engage in abstract and mathematical reasoning – Piaget’s constructivist theory dominated the field for a time, positing that children are born as blank slates and only acquire a sense of number and logic after several years of interaction with the world. A famous example of what led Piaget to believe that children do not possess an abstract concept of number until a certain age is their apparent failure to the “number conservation task” before 6-8 years old. In this task, children are shown two equal rows of checkers, equally spaced, and asked if one of the two rows has more checkers, or if they have the same number. Children usually reply that “it’s the same thing”. Then, while the child is watching, the experimenter increases the spacing of the checkers in one of the two rows and asks the same question again. Children younger than 6 tend to fail and reply that the row in which the spaces between the checkers have increased (and thus in which the total row’s width is wider) has “more”. This was interpreted as evidence that children did not possess any kind of mental representation of numbers before a certain age.

However, it has since been shown that children’s failure on this task did not originate from a lack of understanding of the concept of number, but rather from their interpretation of the questions they were being asked. In 1967, Mehler and Bever notably showed that replacing the checkers by candies and asking 2 to 4 years old which row they wanted to eat led a majority of them to pick the row with the highest number of candies, regardless of the row’s actual width. Thus, by providing sufficient motivation and sidestepping language comprehension difficulties, 2-year-olds could actually show an understanding of number that Piaget thought they did not possess before their 6th year at the earliest. Thus, not only does adult reasoning fail to follow the laws of logic in specific situations, but children are actually capable of manipulating abstract concepts such as numbers when asked the right questions. The Aristotelian view that children are irrational beings that become completely fluent in logical inferences once reached adulthood has long been discarded, in favor of a more nuanced approach accounting for reasoning biases and the development of logical thinking. But to what extent is human reasoning constrained by the content on

which it operates? The question is worth asking, given the great variety of situations in which adults are expected to reason and behave consistently.

Content shaping thoughts and problem solving

According to Evans (1991), the issue of context is one of the crucial questions that any theory of human reasoning needs to address. As we mentioned at the beginning of this introduction, context variations can influence one's ability to engage in logical inferences (e.g. Bonnefon & Villejoubert, 2007; Byrne, 1989; Chao & Cheng, 2000; Cummins, Lubart, Alksnis, & Rist, 1991; Daniel & Klaczynski, 2006; De Neys, Schaeken, & D'ydewalle, 2002; Douven, Elqayam, Singmann, & van Wijnbergen-Huitink, 2018; Evans, Barston & Pollard, 1983; Johnson-Laird, 2006; Quinn & Markovits, 1998; Thompson, 1994, to name a few of the numerous studies on propositional reasoning). However, there is more to human reasoning than making logical inferences from clearly stated propositional sentences, and valuable insights into the influence of context can be found in the neighboring field of problem solving.

One of the perks of problem solving is that it makes it possible to investigate reasoning processes in a wide diversity of situations, closer to real-life concerns than abstract logic problems – albeit, arguably, still far removed from truly ecological settings. The study of problem solving first gained traction around the beginning of the 20th century, among behaviorist researchers interested in animals' problem solving skills (Hull, 1943; Thorndike, 1898; Watson, 1930). They proposed that learning how to solve problems was the result of a progressive process based on trials and errors. Through conditioning, they argued that the animals found the correct solution to puzzle problems by chance, and then reinforced the behavior that led to it until it was automatized. This view was promptly criticized for its inability to account for “intelligent” problem solving strategies in which intermediate goals need to be set in order to find the solution (Köhler, 1917, 1925).

By contrast, tenants of the Gestalt Psychology argued that problem solving was a productive process in which restructurations of the problem's representation (*insights*) may happen and lead to a sudden certainty of a correct response (Mayer,

1983; Wertheimer, 1959). A benefit of this approach was that it could account for the negative impact that past experiences may have on the ability to find the solution to a new problem (e.g. Katona, 1940; Luchins, 1939, 1942). Gestalt psychology notably led to the discovery of functional fixedness, a cognitive bias restraining one's ability to perceive unusual uses of common objects. In a famous study, Duncker (1945) designed different tasks to evaluate how often participants could re-purpose an object they were provided with to solve a problem. In his *Candle Problem*, he asked participants to attach a candle to the wall so that it did not drip onto the table below, using only a box of tacks, matches, and the candle itself. Only few participants managed to find the solution consisting in repurposing the empty box of tacks as a candleholder and tacking it to the wall (see Fig. 1).

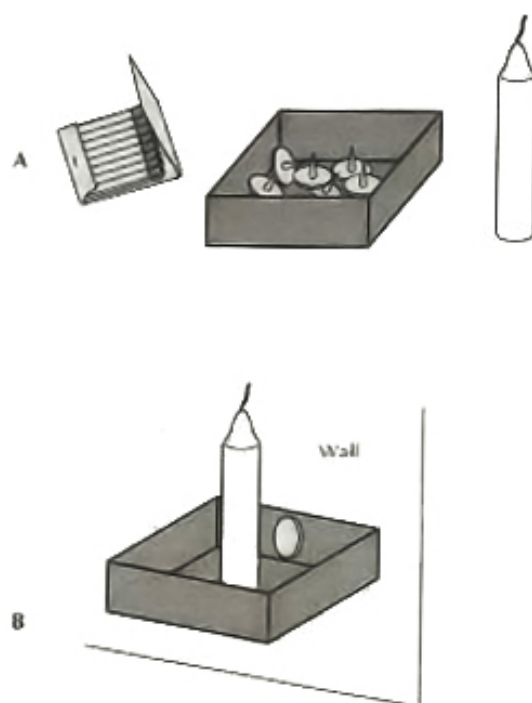


Fig. 1 Initial state (A) and Final state (B) of the Candle Problem. Reprinted from Duncker (1945)

However, when participants were presented with a pile of tacks and an empty box separately, they were significantly more likely to use the box as a support for the candle. In other words, participants experienced difficulties when the box was assigned an initial function that did not fit with the problem (container of tacks), but not when the box was presented on its own. This experiment is a striking example of the influence that context, in the form of prior information about an object's intended use, can have on one's ability to reason in a given situation.

A crucial aspect of the Gestalt's contribution to the description of problem solving resides in its introduction of the notion of an internal problem representation, which can be influenced by prior knowledge and experiences. However, this approach has been criticized for being somewhat underspecified, as well as for saying little about the processes behind the phenomenon of *insight*, thus making it hard to empirically evaluate (Clément, 2009). Addressing these issues, the notion of “problem space” arose from Newell & Simon's (1972) work, in which they defined problem solving as the exploration of a problem space constituted of all the possible states and transitions between states of the problem, according to the solver's interpretation. This idea that problem solving depends on one's interpretation of the problem is a crucial step forward in accounting for performance variations between problems sharing the same structure but differing in their surface elements (Kotovsky, Hayes, & Simon, 1985; Simon & Newell, 1971). Consider, for instance, the famous Tower of Hanoi Problem, defined as follows (Pretz, Naples, & Sternberg, 2003, p. 7):

There are three discs of unequal sizes, positioned on the leftmost of three pegs, such that the largest disc is at the bottom, the middle-size disc is in the middle, and the smallest disc is on the top. Your task is to transfer all three discs to the rightmost peg, using the middle peg as a stationing area, as needed. You may move only one disc at a time, and you may never move a larger disc on top of a smaller disc.

The complete problem space of this problem includes all the possible transitions from the Initial State to the Goal State and allows the use of a 7-step solving strategy (Fig. 2, A.). However, some solvers (notably 6 and 7 years old) add an unnecessary constraint in their interpretation of the problem, by thinking that only movements between two neighboring pegs are allowed (Richard, Poitrenaud, & Tijus, 1993). Thus, these children construct a truncated problem space with a “no peg jumping” rule, which considerably increases the number of steps (26) to find the solution (Fig. 2, B). Interestingly, Clément and Richard (1997) showed that it was possible to lead even adults to construct a truncated problem space with a “no jump” rule by modifying the problem's surface features. For instance, if instead of disks moving between pegs, the problem is stated in terms of individuals moving between floors using an elevator, then participants are more likely to

believe that movements are only allowed between nearby floors since they understand that an elevator does not usually “jump” floors.

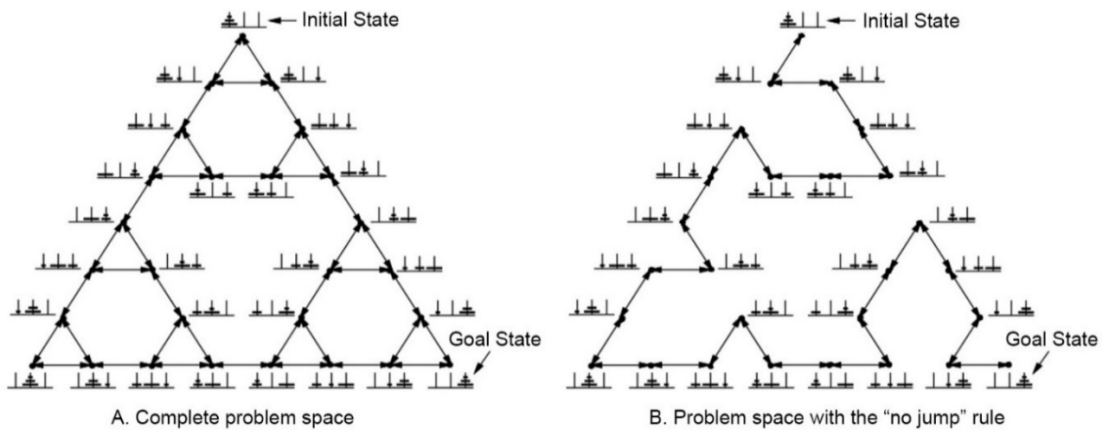


Fig. 2 Complete and truncated problem spaces of the Hanoi Tower problem

By putting emphasis on solvers' interpretations of the situations, these studies on problem solving provided key insights into the part played by everyday knowledge in adult reasoning. This idea that general knowledge about the world may influence the representation of a given problem and dictate how someone will tackle its solving is of importance and will be the subject of further development in the first chapter of this thesis. Overall, this very brief account of content effects in problem solving – and in reasoning in general – underlines the key issue of the interactions between domain-independent knowledge and domain-specific knowledge. Despite its ability to conceive of abstract ideas, human reasoning always happens within a context – be it a real-life situation or a lab experiment setting – and so it seems important to strive to understand how contextual information interferes with our thought processes.

Current undertaking

Among the vast field of reasoning, our focus is on mathematical reasoning in particular, and the manner in which human conceive of numerical situations. Indeed, what better avenue to study the impact of concrete features on the realm of abstract ideas than through mathematics, the very champion of abstraction. While it has been suggested that the surface features of a situation often correlate with its deeper principles (Blessing & Ross, 1996; Gentner & Medina, 1998), mathematics remains a domain in which context-independent reasoning is praised, trained, and considered at least partially mastered after a few years of

formal training. However, the thesis we defend concerns the existence of a worldly influence – taking the form of wording interferences – on human mathematical reasoning.

While the idea that mathematics is grounded in nature has gained traction in the past decades, with works highlighting the conceptual metaphors (Lakoff & Núñez, 2000) and tacit models (Fischbein, 1989, 1994) that underlie key mathematical concepts, the role played by general knowledge in adults' representations of mathematical problems remains a debated issue (e.g. Bassok, 2001; Kintsch & Greeno, 1985; Staub & Reusser, 1995). We intend to show that prior, non-mathematical knowledge about the entities described in a word problem can shape its semantic encoding and dictate one's reasoning about the mathematical situation described. The first chapter of this thesis will elaborate on this idea and present the conceptual framework synthesizing our view.

Arguing for the notion that different representations are constructed depending on the everyday knowledge elicited by a problem is an arduous task, since there exist no direct means of investigation of one's mental constructs. Luckily, many indirect routes have been proposed by cognitive scientists seeking to inspect the different shapes in which human thoughts may come. Researchers have long moved past resorting to fallible methods such as introspection or self-report, and a variety of innovative indirect approaches have been designed to examine participants' mental representations. Among them, the study of reaction times (e.g. Rosch, 1975), metaphors (e.g. Lakoff & Núñez, 2000), gestures (e.g. Fuhrman & Boroditsky, 2010), eye movements (e.g. Verschaffel, De Corte, & Pauwels, 1992), growing lines estimation (e.g. Casasanto & Boroditsky, 2008), operand recognition (e.g. Thevenot & Oakhill, 2006), drawings (e.g. Vosniadou & Brewer, 1992), relational priming (e.g. Bassok, Pedigo, & Oskarsson, 2008), looking times (e.g. Izard, Sann, Spelke, & Streri, 2009), classification (e.g. Chi, Feltovich, & Glaser, 1981), inductive projection (Inagaki & Hatano, 1996), written statements (e.g. Pinnegar, Mangelson, Reed, & Groves, 2011), and strategy use (e.g. Clément & Richard, 1997) constitute a representative – albeit far from exhaustive – list. In our work, we set out to gather converging evidence from complementary approaches to shed light on the representations constructed in the course of arithmetic word problem solving. In particular, we intended to show that numerical situations could, depending on the semantics imbued in their

statement, foster one of two encodings: either a cardinal representation, or an ordinal representation. The difference between cardinality and ordinality – as well as what it entails in terms of semantic encoding – will be introduced in detail in the second chapter of this thesis. We believe that this distinction is so pervasive that its effects can be felt in a wide range of activities, including (but not limited to): strategy use (chapters 2 to 7), categorization (chapter 2), comparison (chapter 2), solvability assessment (chapters 2, 4, & 5), response times (chapters 2, 4, & 5), problem recall (chapter 3), sentence recognition (chapter 3), eye movements (chapter 5), pupil dilation (chapter 5), drawing production (chapter 6) and transfer (chapter 7), while also affecting different populations, among which children (chapter 6 & 7), lay adults (chapters 2 to 6), expert mathematicians (chapter 4) and math teachers (chapter 5), in French (chapters 2 to 7) and in English (chapter 3).

The **first chapter** extends the present introduction. It is a theoretical paper introducing the current theories of arithmetic word problem solving. The contributions of the main models in the field are discussed, followed by a new proposal describing our perspective on the issue, in the form of a theoretical model of semantic congruence (SECO). While this chapter intends to clarify the conceptual framework shaping our work in the following chapters, it should be noted that the main goal of this thesis is not to defend SECO, as the experiments reported in chapters 2 to 7 were not specifically designed to evaluate its predictions. Rather, this thesis aims to describe how the interaction between world knowledge and mathematical knowledge shapes the encoding of arithmetic word problems, especially by leading to the construction of cardinal and ordinal representations of the situations. In fact, SECO and the experiments we report in the next chapters were developed simultaneously, and SECO was molded by the experiments we ran as much as it shaped their designs. The first chapter should thus be seen as an extended introduction presenting the view that we came to embrace while investigating arithmetic word problem representations.

The fundamental distinction between cardinal and ordinal representations is introduced in the **second chapter**. This chapter details the theoretical reasons that led us to consider the role played by the difference between cardinality and ordinality. It also presents the materials that will serve as the basis for all of the experiments reported in this thesis: new arithmetic word problems using different

quantities meant to foster either a cardinal encoding or an ordinal encoding. A first categorization experiment is used to validate our choice of materials, and 5 subsequent experiments evaluate the influence that the cardinal versus ordinal distinction holds on several mathematical reasoning tasks.

By investigating the recall and recognition of specific word problems, the **third chapter** intends to go one step further and show that the distinction between the cardinal and ordinal problems created in Chapter 1 has an influence spanning beyond regular problem solving activities. The nature of the representations stored in memory is investigated by targeting specific forms of misremembrance revealing differences in the inferences made while solving the problems. Problem recall and sentence recognition are thus used to probe participants' problem representations.

The pervasiveness of the effects reported in the first chapters is assessed in the **fourth chapter**, by studying expert mathematicians' ability to overcome the difficulties imposed by semantic incongruence. By examining the influence of domain-independent knowledge on experts' performance, this chapter attempts to provide a fuller characterization of the extent to which human reasoning is constrained by its object. This chapter also replicates the effects observed in the 6th experiment of the first chapter among lay adults, in a new, time-constrained setting.

The **fifth chapter** gathers behavioral and eye tracking data to deal with the different kind of expertise that teachers and pre-service teachers have about the peculiar exercise of word problem solving. The two experiments of this chapter follow two complementary objectives: the first experiment aims at showing that even math teachers are not immune to the deleterious influence of domain-independent knowledge on mathematical reasoning. It constitutes a replication of the experiment in Chapter 4. By recording the eye movements of pre-service teachers performing the same task as the teachers in the first experiment, the second experiment seeks to provide a finer-grained understanding of the differences between cardinal and ordinal representations of arithmetic word problems, as well as of what performing a semantic recoding really entails in terms of cognitive load. The study of fixations, backward eye movements and pupil dilation gives an overview of participants' information collection strategies while attempting to solve cardinal and ordinal problems.

The **sixth chapter** introduces a developmental perspective on the issues at the heart of this thesis. By comparing strategies used by 2nd graders, 5th graders, and adults to solve isomorphic problems, we attempt to draw a picture of the development of cardinal and ordinal representations throughout the years. We study which solving strategies are preferred at each age and illustrate the life-long influence that domain-independent knowledge has on mathematical reasoning. Additionally, a drawing production task provides new insights into the differences between the representation of cardinal and ordinal problems. This chapter also introduces a different, simplified version of the materials used in previous chapters, designed to help 3rd graders understand the mathematical structure of the problems.

Finally, the **seventh chapter** briefly illustrates the negative influence that semantic incongruence may have on fifth and sixth graders' ability to apply a presented algorithm to new problems. This chapter focuses on negative transfer between worked-out examples and new problems sharing the same solution principle, but differing to varying degrees from the training problems. Using new materials differing slightly from the ones used in chapters 2 to 6, this last experiment opens the question of what can be done to foster semantic recoding and improve transfer between problems sharing the same solution principle. This issue, crucial for mathematics education, will be the focus of further consideration in the general discussion of this thesis.

Overall, 16 experiments are presented, conducted among a total of 2180 participants from different backgrounds. In view of the present replicability crisis in experimental psychology (Open Science Collaboration, 2015), we attempted to provide both exact and conceptual replication throughout the different experiments in this thesis. The goal was to collect converging evidence following different approaches in order to assess the validity of the hypothesis that general, non-mathematical world knowledge may foster one of two semantic encodings and interfere with arithmetic reasoning. Although all of the experiments we present build on similar materials, variations are introduced between experiments and between chapters to maximize the reliability of our findings. We hope that the diversity of issues, methods, populations and analyses presented in these pages will engage our readers' attention and spark their interest as ours has been by this fascinating topic.

References

- Aristotle, R. (1999). *Nicomachean Ethics*. In T. Irwin (2nd Eds.), *Nicomachean Ethics*. Indianapolis, IN: Hackett Pub.
- Bassok, M. (2001). Semantic alignments in mathematical word problems. In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 401–433). Cambridge, MA: MIT Press.
- Bassok, M., Pedigo, S. F., & Oskarsson, A. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 343–352.
- Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3), 792.
- Bonnefon, J. F., & Villejoubert, G. (2007). Modus Tollens, Modus Shmollens: Contrapositive reasoning and the pragmatics of negation. *Thinking & reasoning*, 13(2), 207-222.
- Boole, G. (1854). *An investigation of the laws of thought: on which are founded the mathematical theories of logic and probabilities*. New York, NY: Dover Publications.
- Byrne, R. M. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31(1), 61-83.
- Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, 106(2), 579-593.
- Chambliss, J. J. (1982). Aristotle's Conception of Childhood and the Poliscraft. *Educational Studies*, 13(1), 33-43.
- Chao, S. J., & Cheng, P. W. (2000). The emergence of inferential rules: The use of pragmatic reasoning schemas by preschoolers. *Cognitive Development*, 15(1), 39-62.
- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive psychology*, 17(4), 391-416.
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive science*, 5(2), 121-152.
- Clément, E. (2009). *La résolution de problèmes*. Paris, FR : Armand Colin.

- Clément, E. & Richard, J.-F. (1997). Knowledge of domain effects in problem representation: The case of Tower of Hanoi isomorphs. *Thinking & Reasoning*, 3(2), 133-157.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31(3), 187-276.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barlow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*, (pp. 163-228). Oxford, UK: Oxford University Press.
- Cox, J. R., & Griggs, R. A. (1982). The effects of experience on performance in Wason's selection task. *Memory & Cognition*, 10(5), 496-502.
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & cognition*, 19(3), 274-282.
- Daniel, D. B., & Klaczynski, P. A. (2006). Developmental and individual differences in conditional reasoning: Effects of logic instructions and alternative antecedents. *Child Development*, 77(2), 339-354.
- De Neys, W., Schaeken, W., & D'ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & cognition*, 30(6), 908-920.
- Descartes, R. (1984). Meditations on first philosophy. In J. Cottingham, R. Stoothoff, and D. Murdoch (Eds.), *The Philosophical Writings of René Descartes*, Volume 2. Cambridge: Cambridge University Press.
- Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2018). Conditionals and inferential connections: A hypothetical inferential theory. *Cognitive psychology*, 101, 50-81.
- Duncker, K. (1945). On solving problems. *Psychological Monographs*, 58(5), 270.
- Evans, J. S. B., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & cognition*, 11(3), 295-306.
- Fischbein, E. (1989). Tacit models and mathematical reasoning. *For the learning of mathematics*, 9(2), 9-14.
- Fischbein, E. (1994). The interaction between the formal, the algorithmic and the intuitive components in a mathematical activity, In Biehler, R., Scholz, R.,

- Strässer, R., & Winkelmann, B. (Eds), *Didactics of mathematics as a scientific discipline*, (pp. 231-245).
- Flavell, J. H. (1963). *The university series in psychology. The developmental psychology of Jean Piaget*. Princeton, NJ, US: D Van Nostrand.
- Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, 34(8), 1430-1451.
- Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 65(2-3), 263-297.
- Giroto, V., Kimmelmeier, M., Sperber, D., & Van der Henst, J. B. (2001). Inept reasoners or pragmatic virtuosos? Relevance and the deontic selection task. *Cognition*, 81(2), B69-B76.
- Hilton, D. J., Kimmelmeier, M., & Bonnefon, J. F. (2005). Putting ifs to work: goal-based relevance in conditional directives. *Journal of Experimental Psychology: General*, 134(3), 388.
- Hull, C. L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*. Boston, MA: Appleton Century.
- Inagaki, K., & Hatano, G. (1996). Young children's recognition of commonalities between animals and plants. *Child development*, 67(6), 2823-2840.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures* (Vol. 22). Psychology Press.
- Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, 106(25), 10382-10385.
- Johnson-Laird, P. N. (2006). *How we reason*. Oxford, UK: Oxford University Press.
- Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British journal of Psychology*, 63(3), 395-400.
- Katona, G. 1940. *Organizing and Memorizing: Studies in the Psychology of Learning and Teaching*. New York, NY: Columbia University Press.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological review*, 92(1), 109.

- Klauer, K. C., Stahl, C., & Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 680.
- Köhler, W. (1917). *Intelligenzprüfungen an Menschenaffen* [Intelligence tests on anthropoid apes]. Berlin: Preussische Akademie der Wissenschaft.
- Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive psychology*, *17*(2), 248-294.
- Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY : Basic Books.
- Luchins, A. S. (1939). *The Einstellung Effect in learning by repetition*, Unpublished doctoral dissertation, New York University.
- Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs*, *54*, 248.
- Manktelow, K. I., & Over, D. E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition*, *39*(2), 85-105.
- Mayer, R. E. (1992). *Thinking, problem solving, cognition*. New York, NY: Freeman.
- Mehler, J., & Bever, T. G. (1967). Cognitive capacity of very young children. *Science*, *158*(3797), 141-142.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251).
- Pinnegar, S., Mangelson, J., Reed, M., & Groves, S. (2011). Exploring preservice teachers' metaphor plotlines. *Teaching and Teacher Education*, *27*(3), 639-647.
- Politzer, G., & Nguyen-Xuan, A. (1992). Reasoning about conditional promises and warnings: Darwinian algorithms, mental models, relevance judgements or pragmatic schemas?. *The Quarterly Journal of Experimental Psychology*, *44*(3), 401-421.
- Pretz, J. E., Naples, A. J., & Sternberg, R. J. (2003). Recognizing, defining, and representing problems. *The Psychology of problem solving*, *30*(3).
- Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition*, *68*(3), B93-B101.

- Ragni, M., Kola, I., & Johnson-Laird, P. N. (2018). On selecting evidence to test hypotheses: A theory of selection tasks. *Psychological Bulletin*, 144(8), 779-796.
- Richard, J.-F., Poitrenaud, S., & Tijus, C. (1993). Problem-solving restructuring: Elimination of implicit constraints. *Cognitive Science*, 17(4), 497-529.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192.
- Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, 26(2), 145.
- Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Psychology Press.
- Staub, F. C., & Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension* (pp. 285–305). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stenning, K., & Van Lambalgen, M. (2001). Semantics as a foundation for psychology: A case study of Wason's selection task. *Journal of Logic, Language and Information*, 10(3), 273-317.
- 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.
- Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & cognition*, 22(6), 742-758.
- Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Monographs: General and Applied*, 2(4), 1-109.
- Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology*, 84(1), 85.
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24(4), 535-585.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly journal of experimental psychology*, 12(3), 129-140.
- Watson, J. B. (1930). *Behaviorism (Revised Edition)*. Chicago, IL: University of Chicago Press.

- Wason, P. C. (1968). Reasoning about a rule. *Quarterly journal of experimental psychology*, 20(3), 273-281.
- Wertheimer, M. (1959) *Productive Thinking (enlarged ed.)*. Chicago, IL: University of Chicago Press.
- Wolfgang, K., & Köhler, W. (1925). *The Mentality of Apes*. New York, NY: Kegan Paul, Trench, Trubner & Co.

CHAPTER 1 – SEMANTIC CONGRUENCE IN ARITHMETIC: A NEW CONCEPTUAL MODEL FOR WORD PROBLEM SOLVING



Presentation

The first chapter of this thesis is a theoretical article introducing a new conceptual framework to understand the interpretative processes at play in the encoding, recoding, and solving of arithmetic word problems. The article presents the main existing models of arithmetic word problem solving and draws on their limitations to propose a new theoretical model (SECO) accounting for the interactions between domain-independent knowledge and mathematical knowledge in the interpretation of the problems.

This chapter serves two purposes. First, it introduces the literature in arithmetic word problem solving that will be relevant for the following chapters. Its intent is not to provide an exhaustive review of the literature, but to highlight specific studies and trends which substantially influenced our work. The major theories accounting for solvers' reasoning process are discussed, and we take a closer look at a selection of 6 experimental works that illustrate key aspects of arithmetic word problem solving.

Second, this chapter also presents the conceptual framework that guided our investigation into the interferences of world semantics on arithmetic word problem solving. The SECO model we describe was developed in parallel to the experiments that are reported in chapters 2 to 7. As a result, it should be noted that these experiments informed SECO's conceptualization as much as SECO influenced their design. With the introduction of SECO, we aim to propose a conceptual lens to apprehend the determinants of arithmetic word problem solving in order to better characterize how non-mathematical world knowledge can interfere with our interpretation of the mathematical situation described in a problem. Before diving into the specific issue of what the perception of cardinality and ordinality entails in our understanding of numerical situations, this first chapter tackles the broader question of what it means to solve an arithmetic word problem.

This article has recently been accepted for publication as is in the journal *Educational Psychologist*.

Abstract

Arithmetic problem solving is a crucial part of mathematics education. However, existing problem solving theories do not fully account for the semantic constraints partaking in the encoding and recoding of arithmetic word problems. In this respect, the limitations of the main existing models in the literature are discussed. We then introduce the Semantic Congruence (SECO) model, a theoretical model depicting how world and mathematical semantics interact in the encoding, recoding and solving of arithmetic word problems. The SECO model's ability to account for emblematic results in educational psychology is scrutinized through six case studies encompassing a wide range of effects observed in previous works. The influence of world semantics on learners' problem representations and solving strategies is put forward, as well as the difficulties arising from semantic incongruence between representations and algorithms. Special attention is given to the recoding of semantically incongruent representations, a crucial step that learners struggle with.

Keywords

cognitive strategies · knowledge representation · learning strategies · mathematics
· learning and teaching · problem solving

Introduction

What does it take to solve an arithmetic word problem? It goes without saying that finding the solution requires to be able to read and understand the problem statement, as well as to handle its numerical values and compute the solving algorithm. But is it enough to simply know how to read and count? Several studies have highlighted robust effects suggesting that solving arithmetic word problems involves processes other than mere procedural ones, that have yet to be accounted for within a unified theory. For instance, Hudson (1983) showed that finding a solution to the problem “There are 5 birds and 3 worms. How many more birds than worms are there?” was considerably more difficult for kindergarteners than answering the question “How many birds won’t get a worm?”, despite striking similarities between these two situations. Bassok, Wu and Olseth (1995) showed that after being taught the algorithmic solution of a problem describing objects assigned to people (*e.g.* computers given to secretaries), participants could more easily transfer it to problems involving objects assigned to people (*e.g.* prizes given to students), rather than to problems involving different semantic relations, such as problems involving symmetrical sets of people (*e.g.* doctors “assigned” to other doctors). In a study with primary school pupils, Coquin-Viennot and Moreau (2003) found that to calculate the number of flowers a florist needs in order to give 5 roses and 7 tulips to each person among 14 people, factorization (*i.e.* adding 5 and 7 before multiplying the total by 14) was more commonly used if the wording mentioned that the flowers were grouped in a bouquet than if it did not. Finally, Thevenot and Oakhill (2005, 2006) showed that the choice between two alternative solving algorithms is influenced by the cognitive costs of each strategy. Facing a problem statement where the solution was usually obtained by calculating the value of “ $x - (y + z)$ ”, they found that participants’ preferences shifted in favor of the more economical sequential strategy “ $(x - y) - z$ ” when presented with higher values.

Separately, these studies have all been accounted for within a given framework of arithmetic word problem solving; either the schema theory (Kintsch & Greeno, 1985), the situation problem model (Reusser, 1990; Staub & Reusser, 1995) or the semantic alignment framework (Bassok, 2001). However, taken together, these studies on wording effects, content effects and re-representation

processes display a range of findings that, to our knowledge, remain to be explained within a common model. To address this issue, we hereby propose a *semantic congruence* – SECO – model accounting for how the interactions between the solver’s knowledge about the world (the world semantics) and the solver’s knowledge about mathematics (the mathematical semantics) mediate the conceptual and procedural sides of arithmetic word problem solving. We believe that such a model should help pave the way towards the development of new instruction methods by providing a unified account of a range of effects whose considerable influence on students at all levels tends to be underestimated. Before further specifying the SECO model’s inner workings, a description of the range of effects that current theories of arithmetic word problem solving do account for seems in order.

Arithmetic word problem solving theories

Numerous works have highlighted the fact that arithmetic word problems which can be solved using identical arithmetic operations may vary greatly in terms of solving difficulty, be they additive (Carpenter & Moser, 1982; Nesher, Greeno, & Riley, 1982; Riley, Greeno, & Heller, 1983) or multiplicative (Greer, 1992; Squire & Bryant, 2002; Vergnaud, 1983) problems. The two most prominent approaches of arithmetic word problem solving which have attempted to account for such effects are the schema and the situation model theories (see Thevenot & Barrouillet, 2015, for a review).

The schema model

The schema model (Kintsch & Greeno, 1985; Rumelhart, 1980; Schank, 1975; Schank & Abelson, 1977) posits that the resolution of arithmetic word problems relies on the creation, activation and implementation of schemata. Schemata are defined as propositional data structures stored in long-term memory, as a result of repeated encounters with problems sharing the same structure. These operator structures, once created, can be activated and implemented with numerical values from any given context (any cover story), thus providing the solver with a valid solving algorithm. According to this view, the solvers read the problem statement and “the verbal input is transformed into a conceptual representation of its meaning, a list of propositions” (Kintsch & Greeno, 1985, p. 111). The solvers then

activate, in their long-term memory, the schema sharing the same propositional structure as the one in the problem statement. They then instantiate this schema with the specific numerical values of the cover story to interpret and solve the problem. For instance, in a compare problem, a sentence such as “Tom has three more marbles than Joe” cues a “have more than” propositional structure which uses three arguments: two corresponding to Tom and Joe’s sets, and one corresponding to the quantitative proposition associated with the comparison (Kintsch & Greeno, 1985). According to Kintsch and Greeno, this propositional structure can be implemented with the values of any problem using a “have more than” proposition and can be used to choose the solving algorithm.

However, the schema theory has been challenged by works showing that minor modifications within the wording of otherwise structurally identical problems led to significant differences in terms of solvers’ performances. Notably, De Corte, Verschaffel and De Win (1985) showed that modifying the wording of problems sharing the same schema impacted both their difficulty and the type of errors solvers make. For example, problems such as “Bob got 2 cookies. Now he has 5 cookies. How many cookies did Bob have in the beginning?” were only solved by 36% of the children in the study, whereas slightly reworded problems such as “Bob had some cookies. He got 2 more cookies. Now he has 5 cookies. How many cookies did Bob have in the beginning?” were solved by 55% of the children.

Another convincing piece of evidence showing the limitations of the schema model was brought by Thevenot and Oakhill (2005), who asked adults to solve problems such as “How many marbles do John and Tom altogether have more than Paul? John has 29 marbles, Tom has 13 marbles and Paul has 26 marbles”. This problem is usually solved with the algorithm $(29 + 13) - 26 = 16$, which could be explained by the schema model by the fact that the word “altogether” activates a Combine schema $(29 + 13)$ and the words “have more than” activates a Comparison schema $(42 - 26)$ (Riley et al., 1983). However, the authors showed that when the numerical values were replaced by 3-digit numbers (e.g. replacing 29, 13, and 26 by 749, 323, and 746, respectively), participants tended to use another algorithm to solve the problem: $(749 - 746) + 323 = 326$. Indeed, since in both cases John has 3 more marbles than Tom, it would be easier to calculate the difference between John’s and Tom’s marbles and add it to the number of marbles

Paul has. Yet, participants only used this strategy when the use of 3-digit numbers made it too difficult to calculate the solution using the other algorithm. This experiment suggests that participants were able to decide not to blindly apply the schemata activated by the problem, and to construct an alternative problem representation instead.

Another argument showing the limitations of the schema theory came from Thevenot (2010), who asked participants to solve arithmetic problems and later presented them with an unexpected recognition task involving problems that were either identical to the source problems, inconsistent with the source problems, or that described the same situations using paraphrases. The results showed that paraphrastic problems had a higher recognition rate than inconsistent problems. Since, in paraphrastic problems, the propositional structure of the initial problems was lost by the paraphrasing, it follows that recognition was not solely based on a propositional representation, contrarily to what the schema view predicts. Thus, additional interpretative processes are believed to come into play and modulate the solver's performance. In this regard, effects of content – interpretative effects linked to the semantic content of the cover stories – have been shown to influence participants' performance in a way that is not accounted for by the schema theory (Coquin-Viennot & Moreau, 2003; De Corte et al., 1985; Gvozdic & Sander, 2017; Reusser, 1988; Thevenot & Oakhill, 2005; Vicente, Orrantia, & Verschaffel, 2007). This significant blindspot in the schema theory explains the need for a more comprehensive model accounting for the content effects reported in the literature.

The situation model approach

Due to these limitations, the schema theory has since lost ground against an alternative approach, which builds on the theoretical frameworks of mental models (Johnson-Laird, 1980, 1983) and situation models (Van Dijk & Kintsch, 1983). This approach originates from Reusser's model, the *Situation Problem Solver* (SPS), which applies the situation model approach to arithmetic word problem solving (1989, 1990, 1993; Staub & Reusser, 1995). The SPS model accounts for the integration by the solver of the set of information present in the problem statement. Namely, it proposes that reading a word problem results in the creation of an episodic situation model featuring every functional relation depicted within the text and presenting an analogous structure to that of the described situation (Reusser, 1990). For example, in Hudson's study (1983) mentioned in the introductory

paragraph of this paper, the “How many more birds than worms are there” problem refers to a static episodic situation model where birds and worms are conceived of as two disjoint sets of entities, whereas the “How many birds won’t get a worm” problem leads to the creation of a dynamic episodic situation model in which the relation between the two sets is highlighted (Staub & Reusser, 1995). The episodic situation model is then translated into a problem model containing the relevant structural elements and relations from the point of view of the question to be answered. This qualitative representation of the problem statement differs from the purely propositional structure proposed by the schema theory. According to Staub and Reusser (1995), this problem model is then reduced to its abstract mathematical gist, which can be translated into a solving algorithm.

Although it builds on the idea that solvers reason based on mental representations analogous to the situations described in the problem statements, the SPS model does not explicitly describe the processes that form those representations. Indeed, according to the situation model view, “the structure of a representation corresponds to the structure of what it represents” (p.18244, Johnson-Laird, 2010). If a perfect structural correspondence is assumed between the representation itself and what is represented from the external world, this means that the former is presumed to be a faithful internalization of an external state. The processes through which this internalization is achieved are not explicitly in the scope of the SPS approach. In particular, the idea that background knowledge of an individual might influence the internalization process and eventually interfere with the faithfulness of the internalization relatively to the external situation is not a significant topic in the SPS model. The notion that the structure of a representation is identical to the structure of what it represents is hardly compatible with the thought that one depicted situation could be interpreted differently by different individuals. In other words, saying that a problem statement is encoded as a representation whose structure is analogous to the problem statement’s is tantamount to saying that only one representation can be encoded from a given problem, regardless of variations in interpretation that can occur over times or individuals.

The semantic alignment contribution

Other works have been more attentive to this issue, showing that solvers’ prior knowledge strongly constrains the representations they construct, in an often

detrimental way (Thevenot, 2017). Bassok et al.'s (1995) showed that the world knowledge regarding the entities involved in arithmetic problems influenced the transfer to isomorphic permutation problems; for instance, problems involving objects and people, such as caddies and golfers, spontaneously evoke an asymmetric structure (“get”), in which golfers are getting caddies and not the opposite since in our world, in most pragmatic contexts, people receive objects, and not the other way around. In contrast, they showed that problems involving two sets of people (*e.g.* kids from two nurseries) evoke a symmetric structure (“pair”), in which children from both nurseries are paired together. These semantic relations between the problem elements thus constrain participants’ representations of the problems. Bassok, Chase and Martin (1998) provided additional evidence for this claim, by giving participants the names of different types of objects and asking them to use these objects to create arithmetic word problems involving either an addition or a division. For objects linked by an asymmetric functional relation (*e.g.*, a container/content relation between vases and tulips), participants created more division problems (*e.g.* the number of tulips divided by the number of vases) than additions. On the other hand, with objects belonging to the same superordinate category, such as tulips and daffodils, participants created mostly additive problems.

This issue is all the more important given how Bassok et al. (1998) showed that the association between subclasses of objects and specific solving strategies is reinforced throughout education by the exercises proposed in mathematical textbooks. They showed that a vast majority of division problems in math textbooks include elements linked by asymmetrical relations whereas additive problems feature elements belonging to categories of the same taxonomic level such as red and blue marbles. This reinforcement throughout the years of arithmetic school teaching may contribute to the development and strengthening of robust solving biases among learners, making it especially important to model these interpretative effects of content to better capitalize on them.

The semantic alignment framework (Bassok, 2001) aims at accounting for these interpretative effects of content. It goes beyond the SPS view by specifying how world knowledge regarding the entities involved in the problem influences its representation by the solvers. It proposes that the solvers’ knowledge about the objects described in the problem cover stories leads them to abstract an *interpreted*

structure. This structure varies from one problem statement to another, depending on the roles defined by the world knowledge regarding the entities, even when those roles are not relevant – or even deleterious – with regard to the mathematical structure of the problems and the task at hand. Thus, the structure that is abstracted from arithmetic problems can facilitate the resolution when the relations it entails are *semantically aligned* with the objective mathematical relations of the problem, that is when the problem’s semantic structure can be used “to infer, by analogy, its objective mathematical structure” (Bassok, 2001, p. 402; Bassok et al., 1998). For example, performing divisions on problems involving oranges and baskets will prove easier than performing divisions on problems involving oranges and apples, because division is semantically aligned with asymmetrical structures such as the one between containers (the baskets) and their content (the oranges). Supporting this view, Bassok, Pedigo and Oskarsson (2008) showed that addition facts are activated when they are primed by categorically related words usually associated with addition (e.g. the pair tulips-daisies is semantically aligned with addition), but not in cases of misalignment, when they are primed by unrelated words and are misaligned with addition (e.g. hens and radios are not usually connected in an addition model). This was confirmed in an ERP study by Guthormsen et al. (2016) who showed N400 and P600 effects indicating a disruption of conceptual integration when participants were presented with misaligned problems (e.g. a problem in which flowers and vases were added together). These results indicate that, in case of semantic alignment, the semantic content of a problem statement can provide crucial clues to the solvers.

Alternative encodings and re-representation

The strength of the previous approaches are their versatility and their ability to each account for a range of effects documented in the literature. However, it seems that one crucial question remains open: how is it possible to solve a problem whose semantic content is misaligned with its solution? How can one ignore those misleading clues and go beyond their initial encoding of a problem statement to reach a solution? Overcoming semantic misalignment would mean abstracting a new, different structure of the depicted situation. For instance, encoding a problem with caddies and golfers as a distributive structure where golfers are assigned to caddies instead of the opposite. However, the issue whether several alternative interpreted structures can be encoded from the same problem statement, by

different individuals or by one individual over time, has yet to be considered. Ross and Bradshaw (1994) showed that the initial interpretation of an ambiguous story could be influenced by the beforehand presentation of another story sharing some degree of similarity with the latter. This suggests that two different semantic structures can be abstracted from a same situation, depending on participants' past experiences and prior knowledge.

Furthermore, studies on re-representation showed that it is possible for the solvers to turn their initial representation into a new one, allowing them to overcome their initial inappropriate interpretation and find the solution (Davidson & Sternberg, 2003; Gamo, Sander, & Richard, 2010; Sander & Richard, 2005; Vicente et al., 2007). For example, to facilitate the solving of a change problem in which a quantity is added or subtracted from an unknown start set, solvers can represent the problem in terms of a part-whole structure and turn it into a search for the unknown part (Riley et al., 1983). Thus, it is important to tackle what precisely happens when a solver's initial encoding of a problem statement fails to trigger the use of an appropriate solving algorithm, and to get a better understanding of how solvers might overcome an earlier inadequate representation and recode the same problem. Bearing this issue in mind, we wish to build on the SPS model and on the notion of interpreted structure in order to provide a unified model addressing the processes involved in arithmetic problem solving.

The semantic congruence (SECO) model

The SECO model is based on the notion of semantic congruence in arithmetic word problem solving, which it defines and operationalizes by accounting for the interactions between world semantics, mathematical semantics, and algorithms. Within the SECO model (Fig. 1), the product of the interaction between world semantics and mathematical semantics needs to be put in correspondence with an algorithm, by means of an interpreted structure.

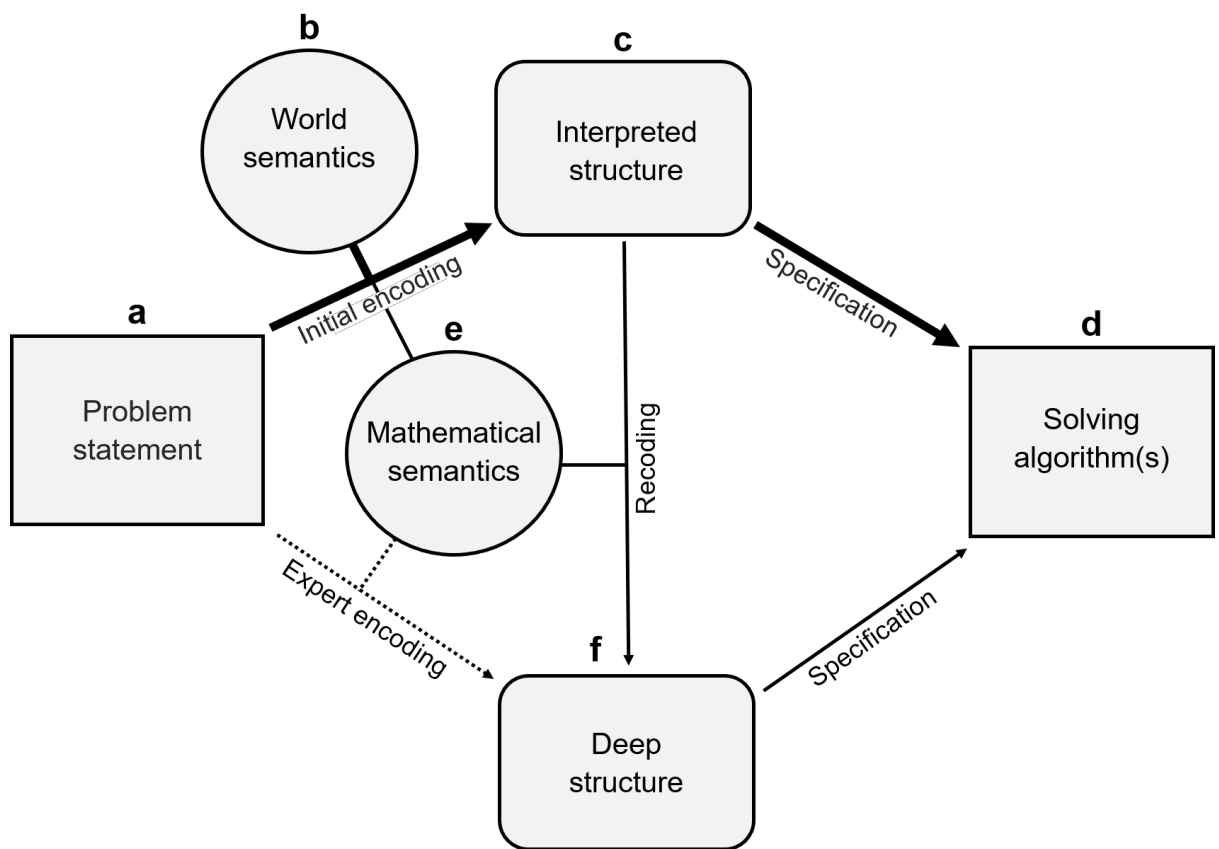


Fig. 1 The SECO model

Components

The components depicted in the SECO model are characterized below; they will be further exemplified in a second phase through six case studies.

- **Problem statement.** The problem statement is a text describing the elements of the problem and the situation(s) in which they interact as well as their relations and associated values.
- **World semantics.** World semantics is characterized by the solver's non-mathematical, daily-life knowledge about the elements of the problem statement as well as the relations between them. For example, world semantics may include knowledge that flowers can be put into vases, that there is a co-hyponym relation between oranges and apples, or that to go from the 1st to the 3rd floor of a building one must pass by the 2nd floor first. There is indeed a broad literature showing that understanding, reasoning, decision-making and problem solving are influenced by the individual's knowledge regarding the entities involved and their relations (*e.g.*, Bassok, 2001; Carey, 2009; Gelman, 2003; Gentner, 1988; Goswami & Brown, 1990;

Johnson-Laird, 1983; Van Dijk & Kintsch, 1983; Kotovsky, Hayes, & Simon, 1985; Stanovich, 1999).

- **Mathematical semantics.** Mathematical semantics is characterized by the solver's mathematical knowledge that is applicable to the problem statement. For example, mathematical semantics may include knowledge that to calculate the size of a set, one needs to add the size of all its subsets, or that to evenly share a collection of objects among several sub-collections, one needs to divide the number of elements in the collection by the number of sub-collections.
- **Interpreted structure.** The interpreted structure is abstracted from the problem statement, integrating pieces of information present in the **text** with the properties, relations and constraints inferred from the world semantics. This notion stems from Bassok and colleagues' research (Bassok & Olseth, 1995; Bassok et al., 1995, see above). Since the mathematical semantics evoked by the problem statement is activated during the encoding, the interpreted structure can feature algebraic values or be instantiated by the numerical values. For example, world semantics about fruits will lead co-hyponyms such as oranges and apples to be encoded as subsets of a superset of fruits.
- **Solving algorithm(s).** A solving algorithm is a finite, unambiguous set of actions that leads to the correct answer when properly executed. Multiple solving algorithms may stem from a given problem statement (*e.g.* De Corte et al., 1985; Gamo et al., 2010; Große & Renkl, 2006; Kouba, 1989; Leikin & Lev, 2007; Thevenot & Oakhill, 2005).
- **Deep structure.** This notion stems from Chi and colleagues' work (Chi, Feltovich, & Glaser, 1981). We define it as the semantic structure integrating the elements of the problem that are relevant for its resolution and describing their relations. This structure does not rely on world semantics but on mathematical semantics. It has been designated as "the objective mathematical structure" (Bassok, 2001), or as "the principle of the problem" (Ross, 1987); for non-mathematical problems the corresponding notion is "the problem space" of an expert solver (Newell & Simon, 1972).

Processes

The processes depicted in the SECO model are characterized as follows:

- **Initial encoding.** This process describes how the problem statement is abstracted into an interpreted structure depending on the world and mathematical semantics evoked by its wording. The world semantics activated by the problem statement constrains the representation of the depicted situation, either by highlighting or by overshadowing specific relations between the problem's entities. Similarly, the mathematical semantics evoked by the problem statement also shapes the mathematical relations represented in the interpreted structure.
- **Specification.** This process describes how an interpreted structure may be specified into an algorithm, as a result of the relations it describes and the numerical values it features. When the relations depicted in the interpreted structure hold a mathematical meaning, they can be translated into relevant operations through this specification process. Not every interpreted structure can be specified into a relevant solving algorithm, since the relations highlighted during the encoding process may not be relevant, and the encoded values may not be the ones needed to solve the problem. A deep structure, on the other hand, may be specified into any relevant algorithm, since it depicts every relevant relation, independently from the influence of world semantics, contrarily to an interpreted structure.
- **Expert encoding.** The expert encoding describes the hypothetical process that may happen when experts initially encode problems within their domain of expertise. As stressed by Chi et al., (1981) experts are believed to be able to disregard the cover story of a problem and directly encode its deep structure. According to this view, an expert may use mathematical semantics and disregard world semantics to directly abstract the deep structure from the problem statement.
- **Recoding.** Since not every deep structure can be specified into a relevant solving algorithm, the recoding describes how, when the initially encoded interpreted structure cannot be translated into an appropriate, tractable algorithm, a new representation can be abstracted by recoding the interpreted structure. This process is akin to the re-representation said to be necessary to overcome difficulties in arithmetic problem solving (Vicente et

al., 2007). It relies on mathematical semantics to recode the situation and build a new structure, closer to the deep structure of the problem. It is a costly process that does not systematically happen.

Inner workings

The SECO model integrates these notions in the following way: it posits that when reading a problem statement (a), the lay solvers will initially encode the problem according to the world semantics (b) as well as to the mathematical semantics (e) evoked by the problem statement, from which they will abstract an interpreted structure (c). This interpreted structure is therefore semantically aligned with the solvers' knowledge about the elements present in the problem statement and can differ from one solver to another for the same problem statement, depending on the state of their world and mathematical semantics. Because it holds a mathematical meaning, this interpreted structure may be specified into an algorithm (d). This algorithm stems from the procedural knowledge that is attached to the mathematical semantics activated by the problem statement. In cases in which no tractable algorithm can be derived from the interpreted structure encoded, the solver faces a dead-end and the need for a recoding process arises. Such a process would appeal to mathematical semantics (e) and not to world semantics, in order to encode a new representation consistent with the deep structure (f) of the problem and thus allow the use of a new algorithm as a result. Contrarily to the interpreted structure from which no tractable algorithm might be derived, this deep structure can be specified into any relevant solving algorithm (d). Finally, the model also introduces the possibility that an individual with sufficient expertise regarding a specific type of problem might directly abstract a deep structure (f) from a problem statement (a), without first extracting an interpreted structure (c) influenced by world semantics (b).

SECO underlines a key aspect of arithmetic word problem solving consisting in the congruence between the semantic knowledge evoked by the problem statement and the mathematical semantics required to find its solving algorithm. If the world semantics attached to the elements in a problem statement is not congruent with the mathematical semantics required to solve the problem, the initial interpreted structure will not be translated into a valid solving algorithm. Indeed, only the mathematical semantics congruent with the world semantics evoked by the problem statement will be used during the initial encoding of the

problem. In cases where the relevant mathematical semantics is not congruent with the world semantics evoked, an extra recoding step is necessary to recode the interpreted structure into a new representation closer from the deep structure of the problem, making the process longer and more difficult.

As in the SPS model, SECO considers that a mental representation of the situation is abstracted when reading an arithmetic word problem. However, contrarily to this model, SECO does not consider that this representation maps onto the structure of the world: by integrating the role of world and mathematical semantics in the encoding of the problem statement, SECO accounts for the fact that there is no unique way to mentally model a problem statement. A situation can be encoded differently by different individuals, and the abstracted structure may be recoded into a new representation if need be.

Accounting for existing results: case studies

In order to better understand SECO's contribution in contrast to the current models of arithmetic problem solving, we propose to tackle representative results in the field through SECO's lens and compare it to the accounts of these results by the two most prominent models of arithmetic word problem solving, the Schema model and the Situation Problem Solver model. As our presentation of SECO shows, its main contribution resides in its depiction of the influence of world semantics on solving strategy choice as well as of the necessity to semantically recode the problems in case of failure. While SECO does not intend to resort solely to world semantics to account for every possible variation in arithmetic problem solving, as other sources of differences exist (*e.g.* algorithm computation abilities or reading comprehension), its central added value consists in its depiction of the influence of world semantics on the encoding, recoding and solving of the problems. We now assess SECO's unique ability to explain the effects reported by a set of six studies mentioned in the introductory section of this paper and presenting representative results in the field. We believe that altogether, these studies prove challenging to the existing models of arithmetic problem solving. We first present two cases illustrating the key issue of the influence of world semantics on the selection of a solving strategy. The following two case studies then showcase the other central feature of the SECO model: its depiction of the existence of a recoding process for semantically incongruent representations. The last two

case studies show how SECO proposes a new take on classical rewording effects, from which important educational implications arise.

World semantics issues

The first two studies we detail illustrate the key influence one's knowledge about the world can have on one's problem solving performance. They describe examples of the effect the content of a problem statement can have on the interpretative processes at play. In other words, they showcase the role of world semantics in arithmetic word problem solving.

Case 1: Bassok et al.'s account of interpreted structures

Empirical findings and authors' perspective. Compelling evidence of the influence of world semantics on the interpreted structure have been provided by Bassok et al. (1995). Participants who were unable to solve an initial permutation problem were presented with a short lesson accompanied by a training problem and its solving equation: $\frac{1}{n(n-1)(n-2)}$ (n being the size of the set of elements being assigned). The participants then had to solve a transfer problem using the same algorithm. The main result was that participants who were first trained on a problem involving an assignment of objects to people ($O \rightarrow P$) had a dramatically higher success rate when they transferred the solution to other $O \rightarrow P$ problems (89% of success) than those who had to transfer the solution to "people assigned to objects" ($P \rightarrow O$) problems (0% of success). According to the authors, the participants interpreted the structure of the problems by using their world knowledge about the roles of the entities involved in the problems, *i.e.* they spontaneously interpreted the problem as a situation in which "objects are given to people" and constructed different interpreted structures depending on which entities were described. The semantic (mis)alignment between the training and transfer problems' interpreted structures accounted for the participants' high (or low) success rate in the transfer problems.

SECO's account of the results. Because it details how an interpreted structure is encoded according to the world and mathematical semantics, the SECO model can account for this result, see Fig. 2 (transfer to objects-to-people problems) and Fig. 3 (transfer to people-to-objects problems). As situations where objects are assigned to people are much more common in daily-life than situations where people are assigned to objects, in SECO, the world semantics (b) regarding the assignment of

elements fosters the idea that objects are usually assigned to people. Therefore, when reading the problem statement (a), the world semantics (b) should, in both “objects to people” and “people to objects” problems, result in an interpreted structure (c) in which objects are assigned to people. This interpreted structure leads the participants to implement the algorithm $\frac{1}{n(n-1)(n-2)}$ (d) with the value corresponding to the size of the set of inanimate objects whereas they should be thinking in terms of which set is being assigned to the other. Indeed, given that participants received limited training, it might be that they did not really understand the solving procedure in the training problem, and thus their mathematical semantics (e) regarding the assignment did not comprise the mathematical notion of “draw without replacement within a set”. Instead, they simply implemented the training algorithm by mapping the semantic roles of the training and transfer problems, and only considered the fact that n was the size of the set of assigned objects in the first problem. Thus, they transfer the algorithm they learnt by replacing the n value by the number of inanimate objects, even if the set of people is the one being assigned to the set of objects. This leads to a correct use of the algorithm in “object to people” transfer problems (Fig. 2) but not in “people to objects” transfer problems (Fig. 3) and accounts for the dramatic contrast between the transfer rates in both conditions (0% vs. 89%).

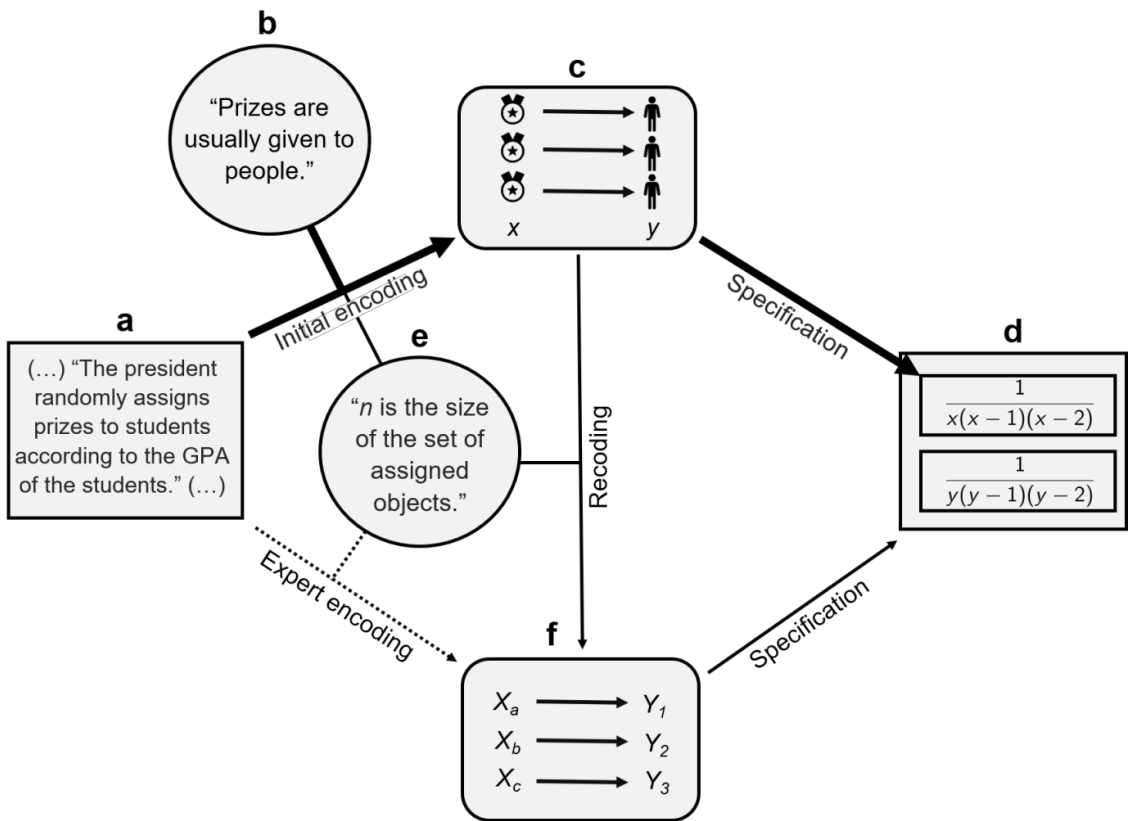


Fig. 2 Modeling of the resolution of a permutation transfer problem with an "objects to people" assignment structure, from Bassok, Wu and Olseth (1995).

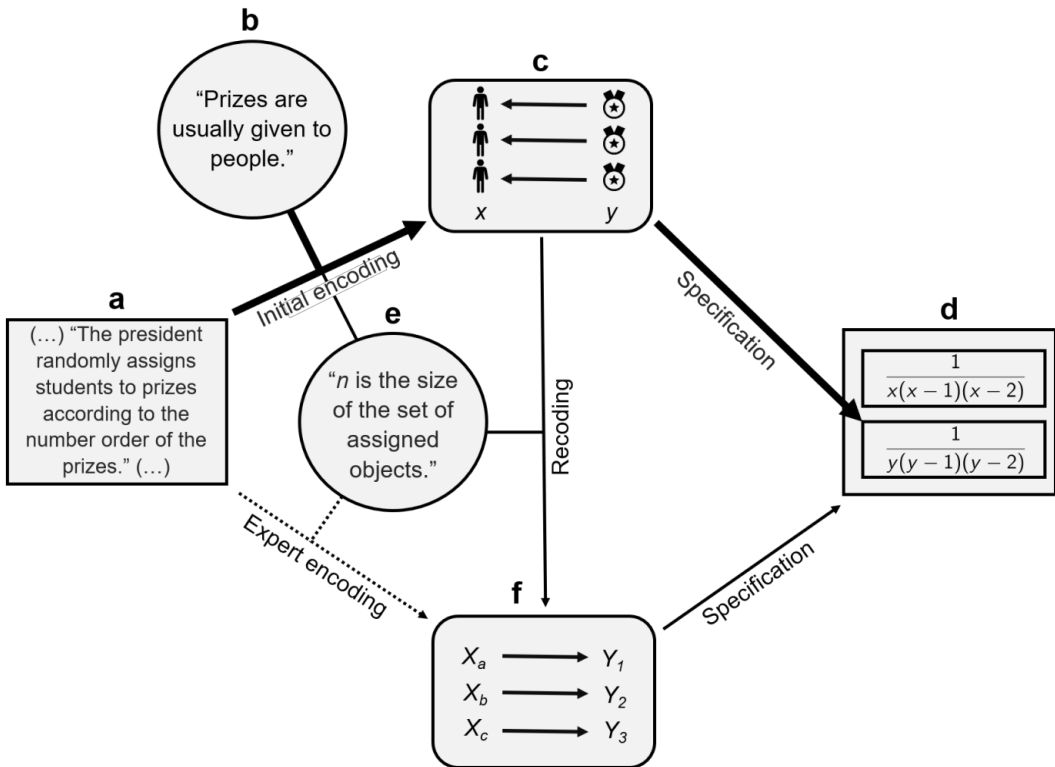


Fig. 3 Modeling of the resolution of a permutation transfer problem with a "people to objects" assignment structure, from Bassok, Wu and Olseth (1995).

Case 2: Coquin-Viennot and Moreau's account of semantic constraints

Empirical findings and authors' perspective. In their study bearing on the use of factorization and expansion algorithms among 3rd and 5th graders, Coquin-Viennot and Moreau (2003) showed that problems such as “For a prize-giving, the florist prepares for each of the 14 candidates 5 roses and 7 tulips. How many flowers does the florist use in total?” were less often solved using factorization (44% among 5th graders) than problems identical in every aspect except for the presence of a superordinate structuring term such as “a bouquet”: “For a prize-giving, the florist prepares for each of the 14 candidates a bouquet made up of 5 roses and 7 tulips.” (68% among 5th graders). This study illustrates how slight modifications in the wording of isomorphic problems can influence the initial encoding. The interpretation proposed by the authors was that the presence of the term “bouquet” favored participants' perception of the two subsets as parts of the same superset and led them to combine the sets into a single entity. We propose a complementary and more systematic explanation using the SECO architecture.

SECO's account of the results. In SECO's view, the use of the word “bouquet” in the problem statement evokes the world semantics stating that a bouquet is a group of flowers, which is compatible with Coquin-Viennot and Moreau's (2003) interpretation. The SECO model would account for these results as depicted in Fig. 4 (problem statement without the “bouquet” term) and Fig. 5 (problem statement with a structuring term).

Since the “no bouquet” problem statement (Fig; 4, a) mentions roses and tulips, the world semantics (b) regarding those elements (*i.e.*, “roses and tulips are two different kinds of flowers”) is activated and favors the encoding of roses and tulips as two disjoint sets in the interpreted structure (c), making the grouping of roses and tulips together less salient. The abstracted interpreted structure (c) thus leads most of the participants to use the expansion algorithm (d) congruent with the representation of tulips and roses as two distinct sets “ $(14 \times 5) + (14 \times 7)$ ”. By contrast, in order to use the factorization algorithm “ $14 \times (5 + 7)$ ”, a solver is either required to infer that tulips and roses can be grouped together (*e.g.* in a bouquet constituted of different flowers), despite the absence of any structuring cue, or to recode the situation (c) according to mathematical semantics (e^{**}) stating that a superset consisting of m sets of x elements and m subsets of y elements has the

same size as a superset consisting of m subsets of “ $x + y$ ” elements, so as to abstract a deep structure (f) of the problem. This deep structure highlights the two different grouping strategies (grouping by individuals or grouping by types of flowers) and is thus congruent with both the factorization algorithm and the expansion algorithm.

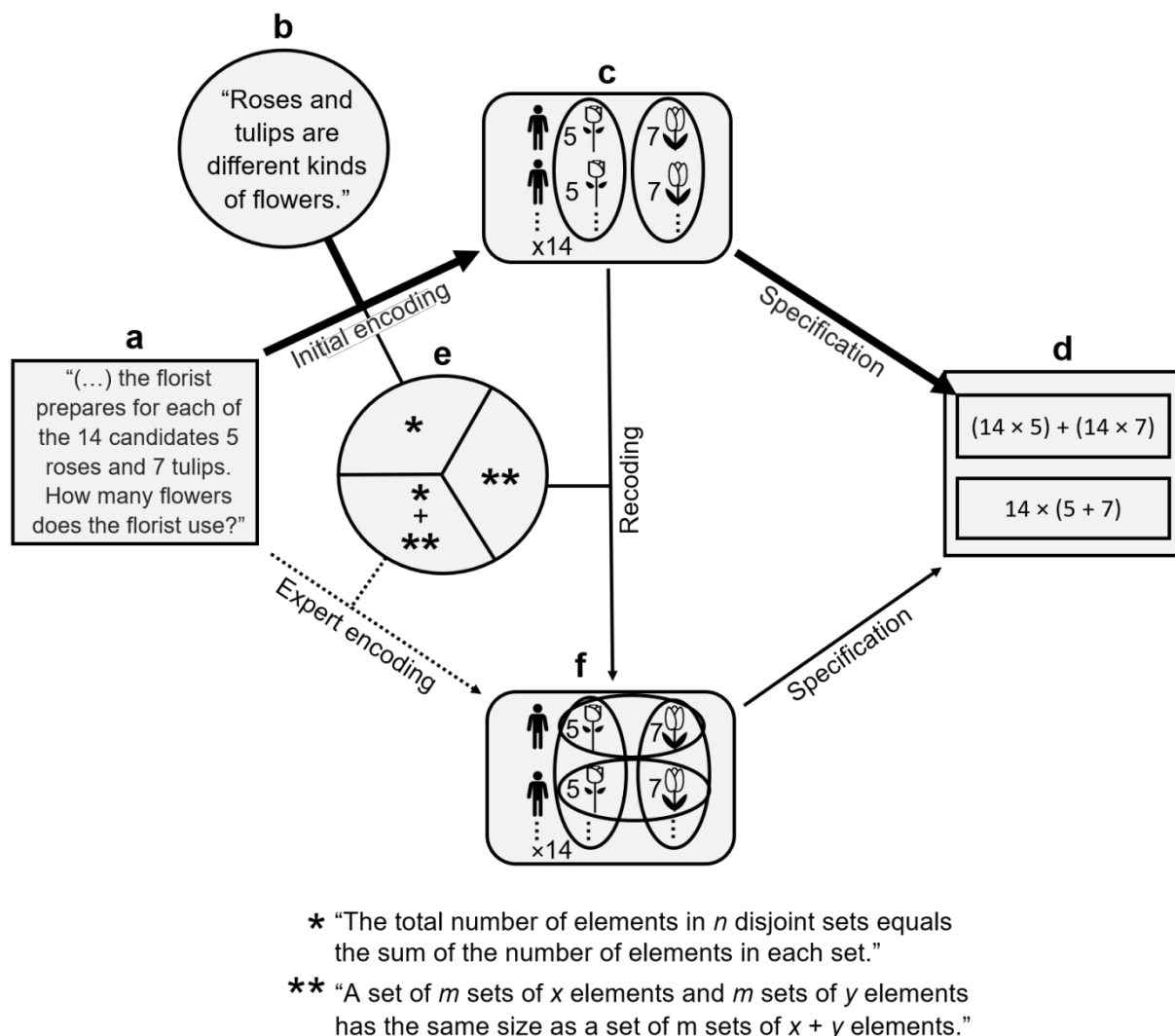
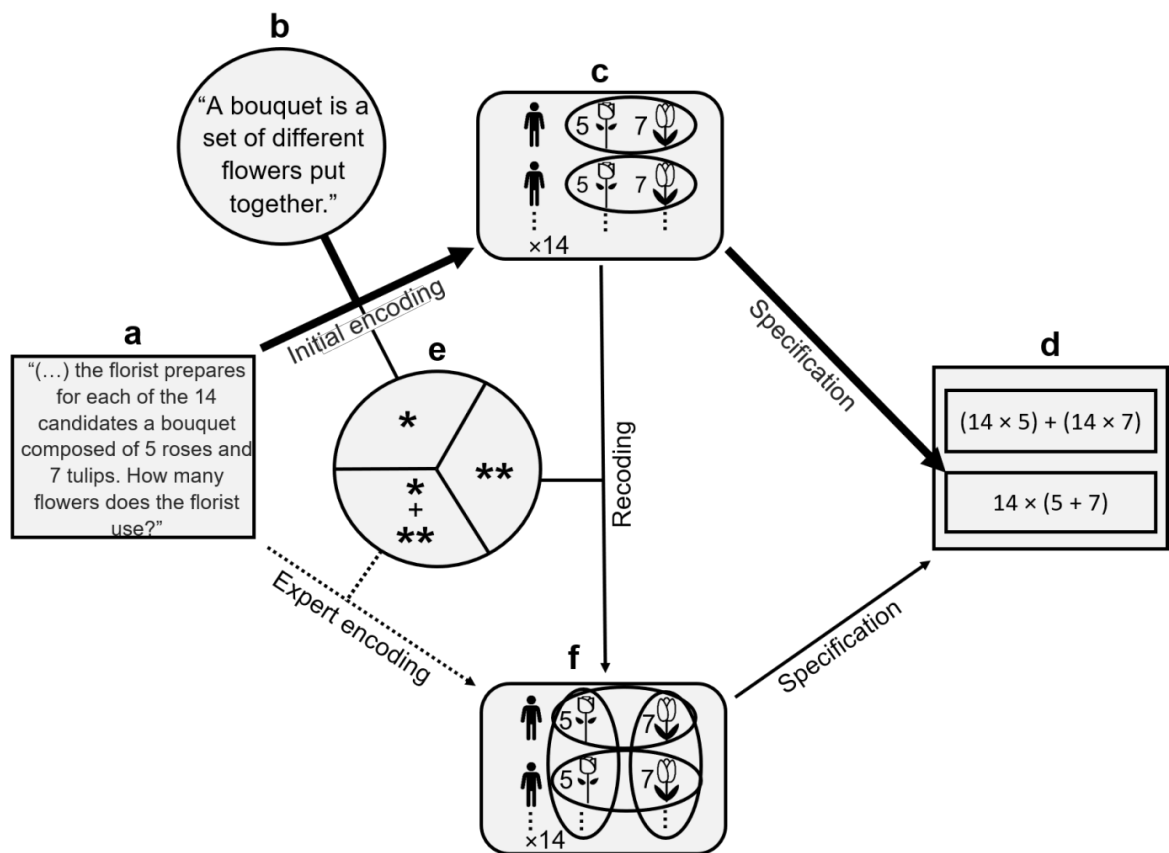


Fig. 4 Modeling of the resolution of a distributive problem without a structuring element from Coquin-Viennot and Moreau (2003).

On the other hand, the resolution of the problem mentioning a structuring element (the bouquet) leads to different steps as detailed in Fig. 5. When the problem statement (a) mentions that the tulips and the roses are grouped together and form a bouquet, then the world semantics (b) related to the bouquet can also be used, in addition to the world semantics related to roses and tulips as flower species. Referring to a bouquet activates the notion of grouping within a single set and

helps the solver encode an interpreted structure (c) increasing the saliency of the two subsets of flowers as parts of the same “bouquet” set, compared to when the structuring element was not mentioned in the wording. The interpreted structure (c) leads the solver to calculate the total number of flowers by adding the number of flowers in each bouquet. Therefore, the factorization strategy $14 \times (5 + 7)$ is the one being mostly used by solvers in this situation. The use of the expansion algorithm “ $(14 \times 5) + (14 \times 7)$ ” is less frequent on such problems and can be the consequence of participants focusing on the distinction between the two types of flowers, roses and tulips, that leads them to count those separately instead of counting the number of flowers within one bouquet first. Alternatively, it can also be the consequence of their explicit use of mathematical semantics (e**) regarding expansion and development.



- * “The total number of elements in n disjoint sets equals the sum of the number of elements in each set.”
- ** “A set of m sets of x elements and m sets of y elements has the same size as a set of m sets of $x + y$ elements.”

Fig. 5 Modeling of the resolution of a distributive problem with a structuring element from Coquin-Viennot and Moreau (2003).

Thus, within the SECO model, the difference between the two problem statements results in a difference between the world semantics evoked by the statements during the encoding process. Different world semantics result in different constraints influencing the encoding of the problems, which lead to different interpreted structures being abstracted, each of them congruent with a specific solving algorithm (expansion when no structuring element is present in the problem statement and factorization otherwise).

Other models' account and their limitations regarding cases 1 and 2

The influence of world semantics displayed by these two case studies is an effect that clearly falls outside the scope of the schema model. In Coquin et al.'s case, there is no theoretically based reason justifying that the addition of the term “bouquet” could influence the selection of a completely different problem schema. Similarly, in Bassok et al.'s (1995) case, the problem schema in the “objects to people” situation should be the same as the one in the “people to object” version, since the only change introduced between the two problems was the semantic nature of the entities constituting the two sets (either people or inanimate objects). In the original schema model, the reader extracts the numerical values and their relations by focusing on the propositional structure of the text (Kintsch & Greeno, 1985). In Bassok et al.'s work (1995), the sentences “The president randomly assigns students to prizes” and “The president randomly assigns prizes to students” have the same propositional structure and should have activated the same schema. However, because one sentence implied that objects were assigned to people, and the other that people were assigned to objects, participants' strategies differed between the two problems. The schema theory alone cannot account for this performance difference without being updated to take into account solvers' knowledge about the problems' entities.

Similarly, the SPS model does not directly integrate the idea that one's knowledge about the entities in a problem could influence the episodic situation model constructed to solve it. Instead, it postulates that the episodic situation model that is built depends on the presentational structure of the problem (text order, narrative point of view, presence of an explicit question, explicitness of relevant relations and so forth) but not on the general knowledge imbued in the problem (Staub & Reusser, 1995). The SPS model relies on the Situational Model assumption that the structure of a representation maps onto the structure of what

it represents (Johnson-Laird, 2010), therefore suggesting that there is only one episodic situation model for each problem, regardless of participants' previous knowledge about the entities featured in a problem. The SECO model, on the other hand, provides a satisfactory account of those results by proposing that the world semantics evoked by a problem also depends on the semantic nature of the elements featured in the problem statement.

Recoding issues

While the first two case studies focused on the mechanisms at play during the initial encoding of a problem statement and their consequences on the solving performances, the next two case studies highlight how an interpreted structure resulting in a dead end can be recoded in certain conditions. In other words, they focus on participants' relative ability to change their initial representation in situations in which multiple mathematical encodings of the same problem statement are possible: different, equally valid representations emphasizing distinct relations.

Case 3: Thevenot and Oakhill's account of alternative representations

Empirical findings and authors' perspective. Studying the influence of number size on the use of solving algorithms, Thevenot and Oakhill (2005) shed light on the factors triggering the recoding of an interpreted structure into a new representation. They investigated the strategies used to solve *compare* problems by using an operand-recognition paradigm consisting in interrupting the presentation of the problem statements to ask participants whether they recognized specific numbers. Recognition performance was used to determine if these numbers were currently maintained in working memory or if they had already been used in calculation and had thus started to fade from memory. They used problems such as “How many marbles does John have more than Tom and Paul together? John has x marbles, Tom has y marbles and Paul has z marbles”. The authors' findings show that participants used the grouping algorithm “ $x - (y + z)$ ” when the task was not especially demanding due to the problem's values being small, whereas they used the more economical sequential strategy “ $(x - y) - z$ ” when the use of larger values implied that the task had higher cognitive costs. Indeed, the second strategy is less cognitively demanding because performing two

successive subtractions allows the solvers to complete a subgoal “ $x - y$ ” while reading the problem, and the result of this operation can be maintained in working memory during the rest of the problem instead of the two initial values (x and y). On the other hand, calculating the value of “ $y + z$ ” and then subtracting it from the value of x requires to maintain the value of x in working memory until the end of the text and the final operation. In other words, if the values 636, 345 and 123 appear in that order in the problem statement, then it is easier to first calculate the value of “ $(636 - 345)$ ” while reading the text and then subtracting 123 from the result later on than to memorize the three values to calculate “ $636 - (345 + 123)$ ” at the end of the problem statement.

SECO’s account of the results. Within the SECO model (see Fig. 6), this effect follows from the fact that the problem statement (a) mentions marbles that are grouped together and then compared. The interpreted structure (c) thus features two disjoint sets: one corresponding to John’s marbles, and the other one to Tom and Paul’s put together. This interpreted structure (c) is semantically congruent with the grouping algorithm “ $x - (y + z)$ ” (d) that is preferentially used for problems with small values. When computing the algorithm becomes impossible because of the larger x , y and z values, some participants need to recode the situation to avoid maintaining the three values in memory. By focusing on the mathematical knowledge regarding parentheses removal (e**), according to which “ $x - (y + z)$ is equivalent to $x - y - z$ ”, participants can recode their initial representation into an alternative representation closer to the deep structure (f) of the problem, in which Tom and Paul’s sets are perceived as two independent sets that can successively be removed from John’s set. They can then switch to the more economical sequential algorithm “ $(x - y) - z$ ” (d). In other words, difficulty to compute the algorithm triggered a re-elaboration process that focused on the mathematical semantics to recode the problem’s representation.

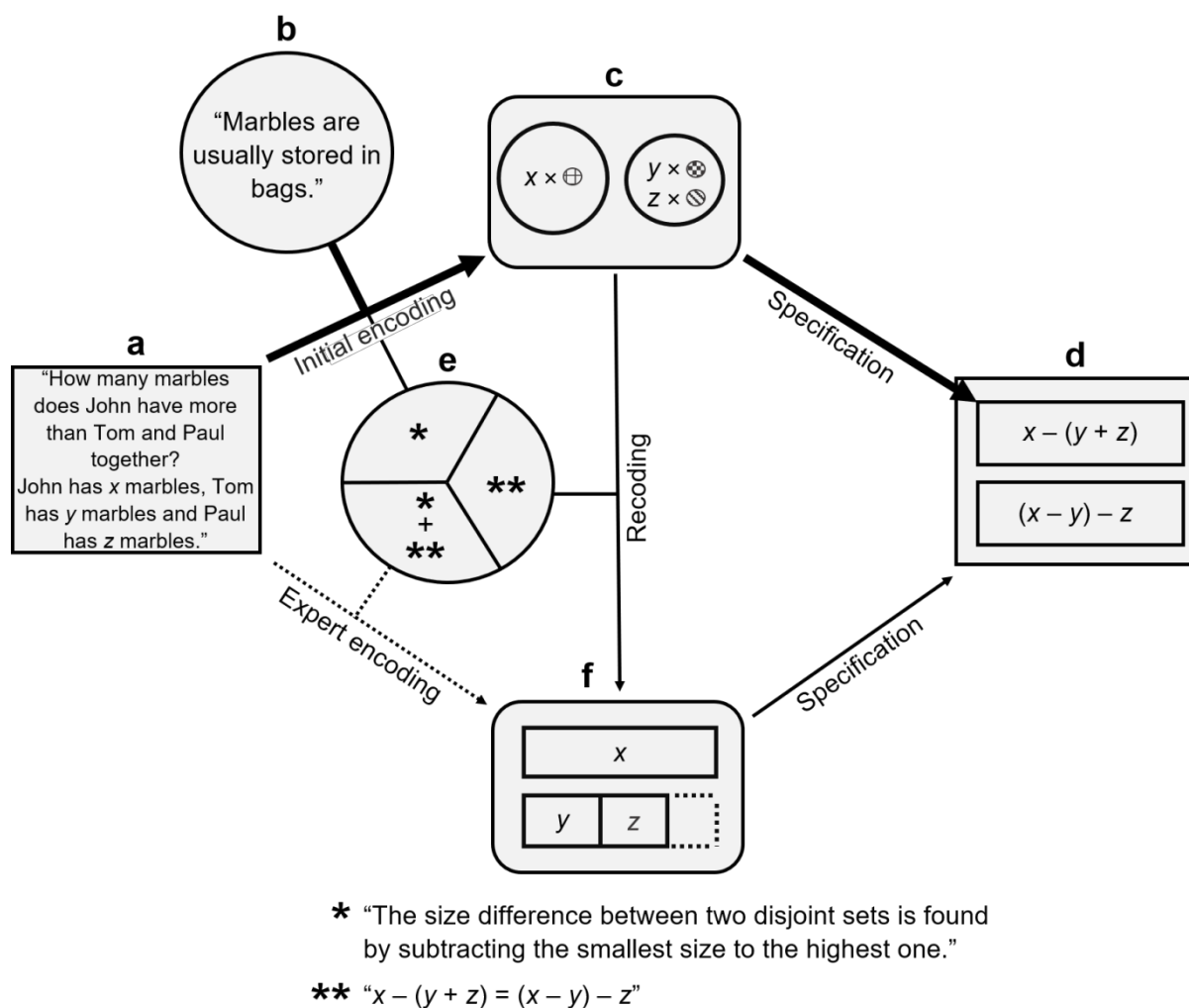


Fig. 6 Modeling of the resolution of a "High Cost" problem from Thevenot & Oakhill (2005). This problem could become either a two-digit problem or a three-digit problem depending on the values given to x , y and z .

Case 4: Gamo et al.'s account of world semantics constraints and semantic recoding

Empirical findings and authors' perspective. In addition to being another illustration of the central role of world semantics on arithmetic word problem solving, the study that Gamo et al. (2010) conducted in 4th and 5th grade classrooms provides valuable insight into the semantic recoding of the initial, inadequate representation of a problem into a new, more polyvalent one. In their study, Gamo et al. used problems that all shared the same formal deep structure, but that involved different types of elements. When the elements were known by the solvers to be unordered entities, such as marbles, scissors or pens, the authors predicted that the participants would abstract an interpreted structure emphasizing the cardinality of the situation, such as an embedded sets structure. This structure

was shown to lead the participants to use a 3-step algorithm to solve the problems. For example, the problem “John bought an 8-Euro exercise book and scissors. He paid 14 Euros. A pen costs 3 Euros less than the exercise book. Paul bought scissors and a pen. How much did he pay?” was preferentially solved using the 3-step algorithm consisting in calculating the price of the pen and the price of the scissors before adding them up: $14 - 8 = 6$; $8 - 3 = 5$; $6 + 5 = 11$. On the other hand, when problems involved ordered units, as is the case in problems involving age, where events are ontologically ordered on the line of time, the authors predicted that the participants would abstract an interpreted structure emphasizing the ordinality of the situation, such as a timeline where different events are represented as positions on an axis. This axis-based interpreted structure would make it possible for the participants to use a different solving algorithm. For example, the problem “Antoine took painting courses at the art school for 8 years and stopped when he was 14 years old. Jean began at the same age as Antoine and took the course for 3 fewer years. At what age did Jean stop?” was predominantly solved using a shorter, more efficient 1-step algorithm: $14 - 3 = 11$. Indeed, the fact that the problem involves durations makes it easy to see that since Jean and Antoine started taking the course at the same age, then the difference between the number of years they each followed the course is equal to the difference between the age at which they stopped taking the course. Thus, the problem can be solved without calculating their age when they started taking the class. Both problems could be solved using both algorithms, but depending on the elements featured in the problems, participants preferentially used one or the other of the two strategies.

In the first experiment of the study, the authors studied the conditions allowing for strategy change. They divided the participants in two groups, both of which had to complete a pre-test and a post-test in which they had to solve similar problems using only one arithmetic operation. Between the two tests, one of the groups followed two 60-minutes training sessions during which the children were instructed to compare the two strategies and incited to see how the 1-step algorithm could be used even on number of element problems. They were explicitly trained to identify their initial semantic representation and they were shown a visual representation of the deep structure of the problems to help them recode their initial encoding of the situation. The other group did not receive such a training. The two main findings were that children did solve problems differently

depending on the world semantics they evoked, and that teaching the children to use both strategies by focusing on the mathematical relations between the entities described and by studying the deep structure of the problems yielded significant result in increasing their ability to use the shorter 1-step algorithm on problems with unordered elements.

SECO's account of the results. These findings are a perfect fit within SECO's framework, since they show both how mathematical and world semantics interact in the encoding of the problem statements into an interpreted structure, and how this interpreted structure then either leads to the use of a semantically congruent solving algorithm or is recoded to allow the use of a semantically incongruent solving strategy. Indeed, in the case of an age problem (see Fig. 7), the world knowledge (b) relating to how time events are usually conceptualized (as transitions between positions on a timeline) is evoked by the problem statement mentioning ages (a). This leads the children to encode an interpreted structure (c) in which the events described are represented along a timeline, which lets them directly compare the ages at which they each stopped attending the classes. This structure can then be specified into the 1-step algorithm (d) congruent with it. On the other hand, when reading a problem with unordered elements (see Fig. 8), Gamo et al. (2010) indicate that the encoding is influenced by the students' knowledge (b) that the elements can be grouped together in any order, and that, for example, the scissors can be indifferently grouped with the pen or with the notebook. The resulting interpreted structure (c) has an embedded set structure that leads the students to calculate the value of each subset (the price of each item). This structure can then only be specified into the 3-step solving strategy (d). In order to use the shorter 1-step strategy, the students needed to use mathematical semantics (e) and recode their representation into a new, more polyvalent one (f). This explains why the only group who increased their performance in using the 1-step algorithm on number of elements problems was the one that followed a training based on the mathematical principle behind the use of the 1-step algorithm and the study of the deep structure.

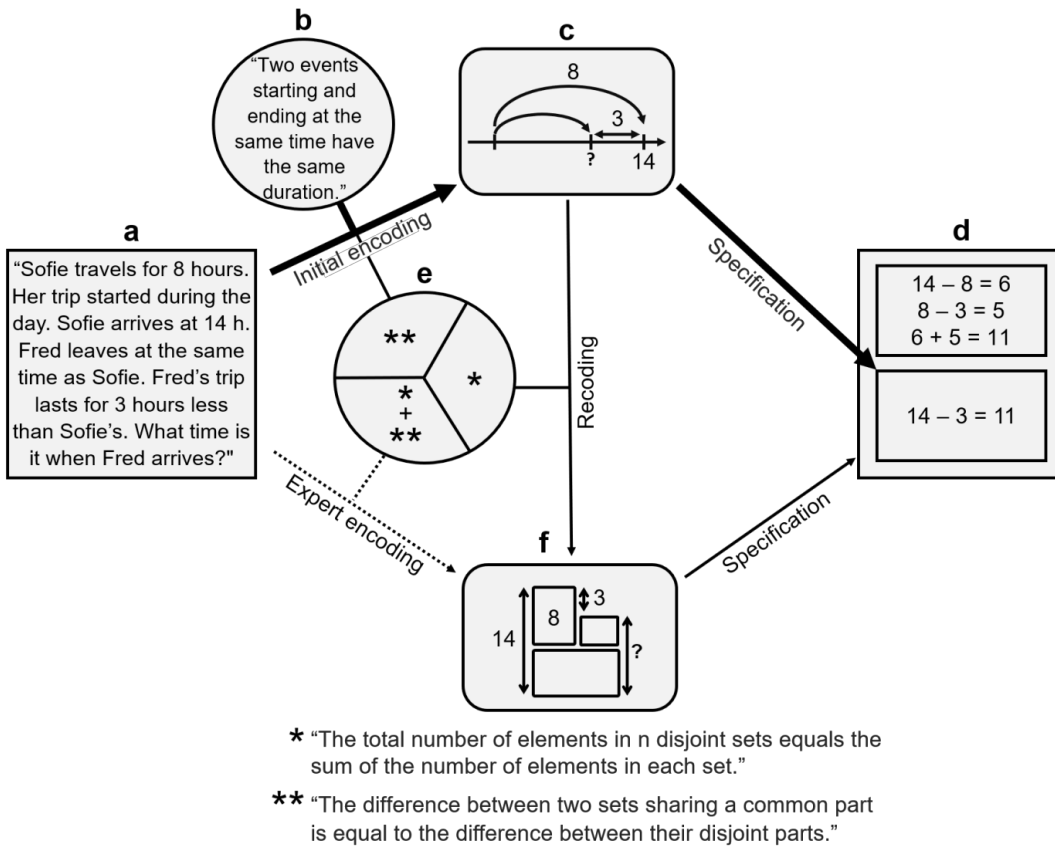


Fig. 7 Modeling of the resolution of an ordinal problem from Gamo et al., 2010.

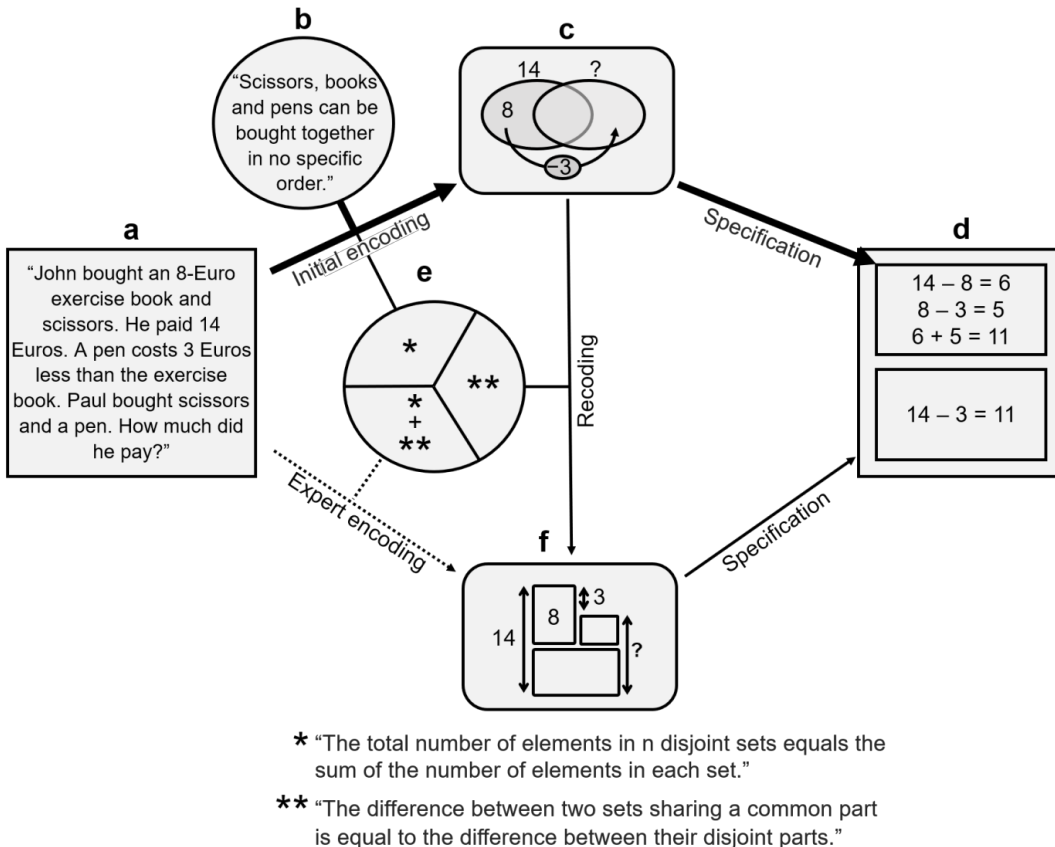


Fig. 8 Modeling of the resolution of a cardinal problem from Gamo et al., 2010.

Other models' account and their limitations regarding cases 3 and 4

These two last case studies showed that when the initial encoding of a problem statement does not lead to a satisfactory solving algorithm, a recoding may happen to encode a new representation congruent with a valid algorithm. As mentioned previously, in Thevenot and Oakhill's case, the idea that a problem could be solved differently depending on whether it features low or high values falls beyond the scope of the schema theory. Indeed, if a schema is constructed from the text-base, then two text-bases differing only by the range of their numerical values should result in two identical schemas being used. Even though it could be argued that students are switching from a schema to another depending on the values provided in the problem statement, such a claim would require a theoretical extension of the schema model accounting for the conditions under which such a switch can occur. Similarly, if the SPS model predicts that one constructs a representation whose structure is that of the described situation, then why would two different representations be constructed based on the same situation? None of the aforementioned models of arithmetic word problem solving directly predicts that an encoding can be recoded depending on how efficient the algorithm it leads to is.

Finally, regarding Gamo et al.'s results, the schema theory may state that some problems correspond to a schema (the so-called ordinal problems) and some do not (the so-called cardinal problems). However, because this theory does not take the structure of the solver's representation into account, it provides no basis to explain why such a schema would only be used on some problem statements and not on others. Specifically, without these semantic features, there is no a priori reason to predict that words such as "age", "during" or "years" would activate a schema corresponding to the 1-step algorithm whereas words such as "scissors", "pen" or "book" would fail to do so. On the other hand, the situation model approach states that a representation analogous to that of the situation described is constructed and used as a basis for reasoning. Because of that, this theory can explain why different problems can be represented differently and thus lead to the use of different algorithms, but the SPS model does not refer to the fact that solvers interpret the situations through the lens of their own previous knowledge. In other words, the situation problem view does not model the constraints imposed by world semantics on the encoding of arithmetic word problems. Interestingly, it can

be noted that the influence of general semantic dimensions such as the cardinal versus ordinal distinction is compatible with the semantic alignment framework. However, in the semantic alignment framework, the question of the recoding of semantically incongruent representations has not been addressed, and SECO's predictions regarding the students' ability to perform a semantic recoding when given appropriate guidance falls beyond this framework. Thus, the fact that the participants were able to solve the number of elements problems using the 1-step algorithm after the training sessions is not predicted by the semantic alignment framework, whereas SECO's take on semantic recoding aided by mathematical semantics offers a reasonable explanation of the effect.

Rewording issues

Several works have highlighted how small modifications in the wording of structurally isomorphic problems could result in significant performance disparities (Cummins, 1991; Cummins, Kintsch, Reusser, & Weimer, 1988; Davis-Dorsey, Ross, & Morrison, 1991; Staub & Reusser, 1992; Stern & Lehrndorfer, 1992; Vicente et al., 2007). Such effects have considerable educational implications, since they illustrate how minor phrasing variations can drastically help (or hinder) the students' understanding of a given problem. As such, they constitute a promising route to assist students in overcoming some of the obstacles they meet in arithmetic word problem solving. Here, we focus on two studies showcasing such rewording effects, to illustrate how SECO can also account for such emblematic results by depicting the changes they entail in the interpreted structures abstracted.

Case 5: Hudson's account of children's understanding of differences between sets

Empirical findings and author's perspective. In his seminal work on numerical differences, Hudson (1983) compared two formulations of comparison problems that led to considerably different levels of performance. Kindergarten children were told there was, for example, "5 birds and 3 worms", and they were asked either "How many more birds than worms are there?" (25% of correct answers among kindergarteners) or "How many birds won't get a worm?" (96% of correct answers among kindergarteners). The author explains that the use of "won't get" reduced the misinterpretation of the "how many more than" construction by highlighting the one-to-one correspondence between the given sets.

SECO's account of the results. The SECO model accounts for those results in the following way. As depicted in Fig. 9, an interpretation of Hudson's findings within the model would be that the sentence "how many more birds than worms are there" in the problem statement (a) evokes aspects of world semantics (b) emphasizing the difference between the two sets of elements (knowledge that birds and worms are two different animal species) thus inducing a comparison between the two groups of elements, without specifying how these two groups should be compared. In contrast, as depicted in Fig. 10, the wording of the problem statement (a) in the "won't get" condition emphasizes the pairing relation between birds and worms and evokes a different aspect of world semantics (b) (*i.e.* "birds eat worms") which promotes the mapping between the two sets within the interpreted structure (c). Thus, in the "more" condition, the interpreted structure (c) consists in two disjoint sets of elements and provides no hint that would trigger a subtraction algorithm.

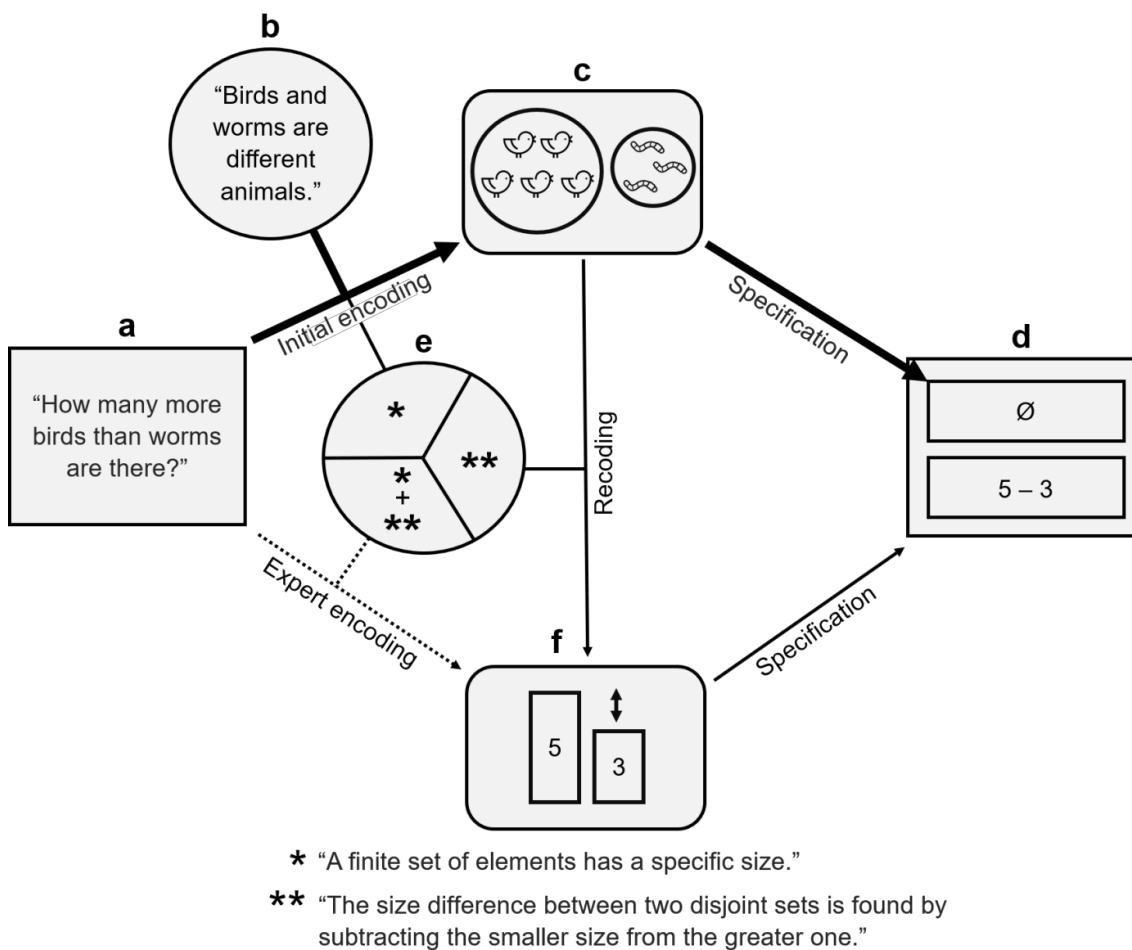


Fig. 9 Modeling of the resolution of a "More" problem from Hudson (1983).

By contrast, the interpreted structure (c) in the “won’t get” condition affords a one-to-one mapping between 3 birds and 3 worms. The “won’t get” condition therefore evokes an interpreted structure that is semantically congruent with an efficient strategy (d), namely counting from 3 to 5. In the “more” condition, recoding the interpreted structure into a deep structure (f) of the problem remains possible, but requires using mathematical semantics (e**) about subtraction, which is not systematically acquired at this early age, thus explaining the low performance on this task (25% among kindergarteners). While Hudson accounted for this finding by stating that comparable constructions of the general form “how many more [...] than?” tended to be misinterpreted, SECO provides an account of this effect in terms of representational differences.

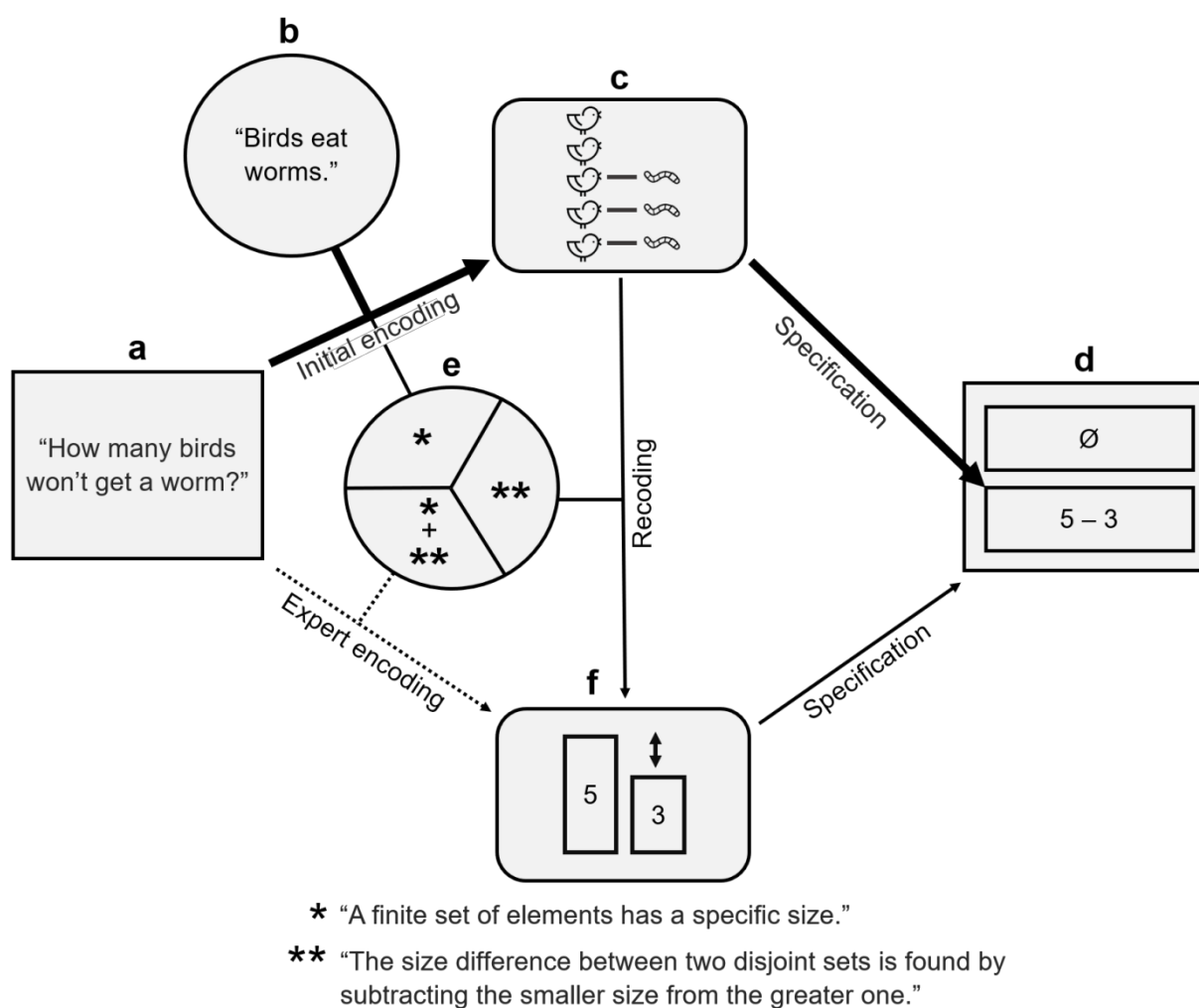


Fig. 10 Modeling of the resolution of a “Won't get” problem from Hudson (1983).

Case 6: De Corte et al.'s (1985) account of rewording effects

Empirical findings and authors' perspective. De Corte et al. (1985) used combine, compare and change problems to study the effects of conceptual rewording on first and second graders' performance, and brought further evidence of the positive effects of specific forms of rewording. For each problem, they compared a "standard" version with a "reworded" version that stated more explicitly the relations between the sets to make them clearer for young students. For example, one of the compare problems they created was "Pete has 8 apples. Ann has 3 apples. How many apples does Pete have more than Ann?". They compared students' performance on this problem and on its reworded version: "There are 8 riders but there are only 3 horses. How many riders won't get a horse?". Results showed that 47% of first graders managed to solve the compare problems in their standard version, whereas 70% of them managed to solve the reworded version. With a rate of success of, respectively, 76% on standard compare problems and 90% on reworded compare problems, second graders also benefitted from the conceptual rewording, although to a lesser extent. The authors explained this difference between the two conditions by stating that only the "won't get" condition provided enough linguistic cues to compute the difference between the sets, whereas the "more" condition remained ambiguous to inexperienced solvers.

SECO's account of the results. SECO provides a complementary account of these results. In the standard version (Fig. 11), the problem statement (a) does not evoke any aspect of world semantics that could help with the matching of the two sets in the interpreted structure (c). Thus, students who have not sufficiently acquired the mathematical semantics (e^{**}) regarding the calculation of the difference between two sets will fail to solve the problem. This explains why standard compare problems had a low rate of success for first graders, and a higher one for second graders.

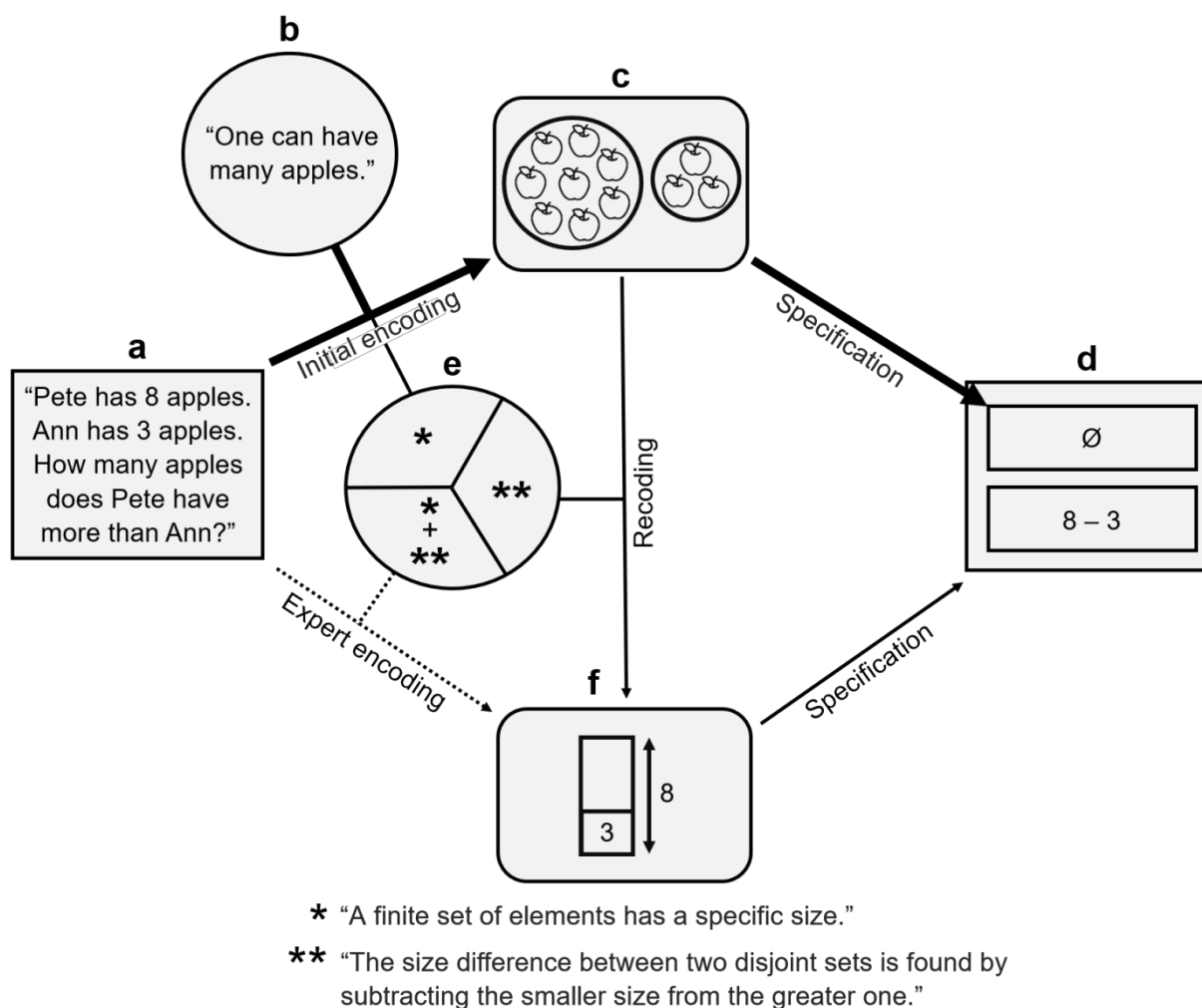


Fig. 11 Modeling of the resolution of a standard compare problem from De Corte et al. (1985).

On the other hand, the reworded problem statement (Fig. 12) evokes knowledge about riders and horses (b) namely the information that a rider is supposed to ride a horse. The interpreted structure (c) thus features the pairing of the three horses with their respective riders and makes it easier to understand how to count the horseless riders remaining. The mathematical semantics (e**) is not necessary in this case to solve the problem, which explains why the performance rate was higher in both age groups.

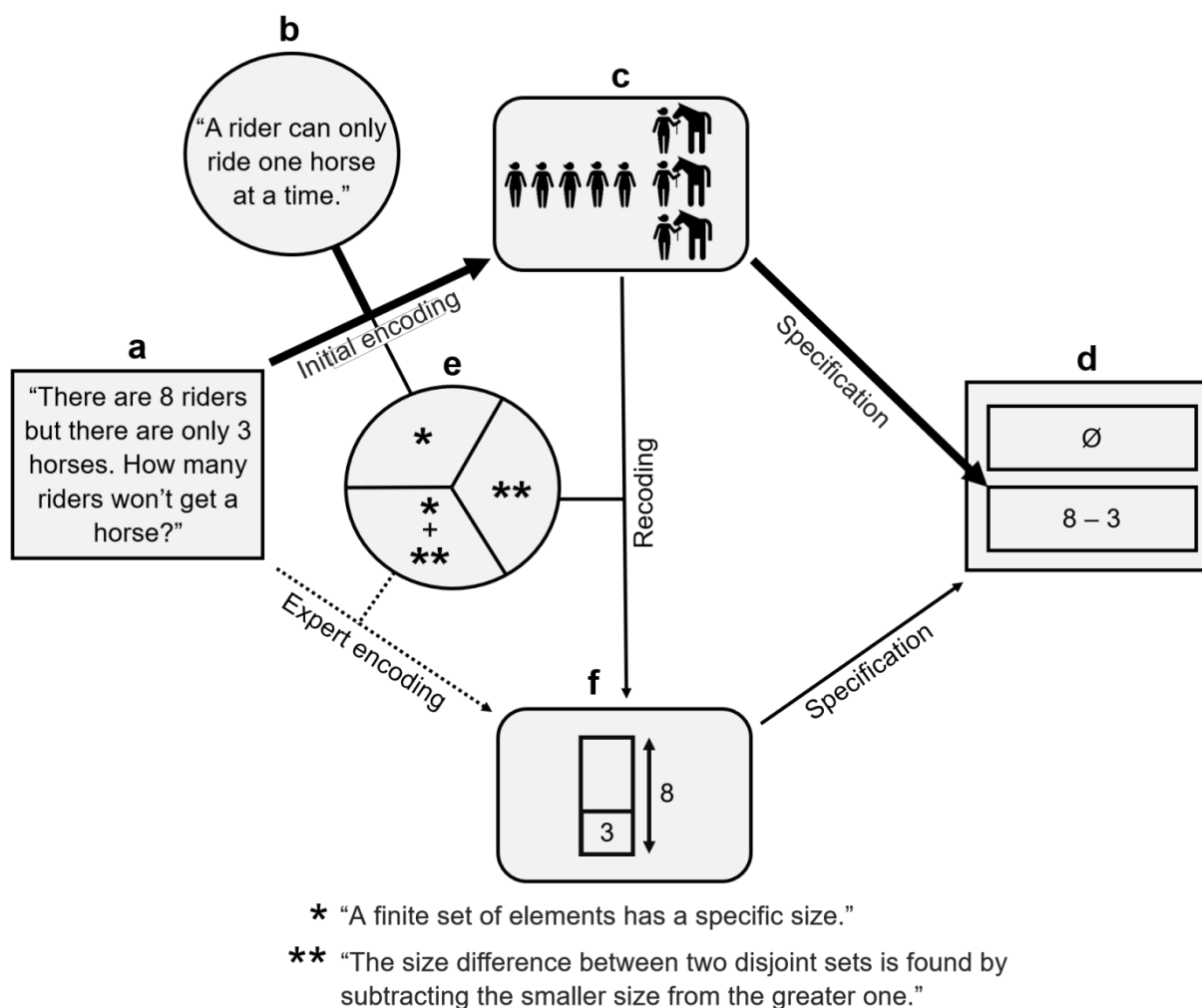


Fig. 12 Modeling of the resolution of a reworded compare problem from De Corte et al. (1985).

Other models' account and their limitations regarding cases 5 and 6

As stated by Vicente et al. (2007), the computational models using problem schema as a basis to explain word problem solving behaviors have struggled to systematically explain the rewording effects of studies such as the two presented above, due to the relatively weak elaboration of the first text-processing stage in their models. Both in Hudson (1983) and in De Corte et al.'s (1985) study, the initial problems and their reworded counterparts shared the same structure according to Riley et al.'s (1983) classification of additive one-step problems. However, small modifications in the wording of the problem statements resulted in significant performance disparities, an effect which the schema model would struggle accounting for.

On the other hand, the SPS model focuses on the idea that a representation, the episodic situation model, is built featuring the relations depicted in the problem statement. According to Staub and Reusser (1995), this representation is different in the two conditions, since the “won’t get” situation imbues the difference with real world meaning, whereas the “more” condition only refers to a static, abstract situation. This suggests that the SPS could have predicted such rewording effects relying on an elaboration of the semantic relations described in the text, since it made the relations between the sets more salient, which explains why the representation was more accurate and led to a higher success rate in the “won’t get” condition. However, it can be noted that any rewording effect capitalizing on prior knowledge, such as replacing computers and secretaries by two sets of doctors in Bassok et al.’s (1995) work, would fall beyond the scope of the SPS model.

Conclusion

When taken together, these six case studies show how SECO can account for varied results within a unified model. While explanations for these results have been provided by one of the already existing theories of arithmetic word problem solving, it appears that none of the aforementioned models can account for all of them simultaneously. In our view, one of SECO’s strengths is that it provides an original integrative framework for the existing results in the literature.

SECO’s added value

The current paper proposes a model detailing the processes at play in arithmetic word problem solving and accounting for how algorithms are found by solvers and how their performances may differ depending on the task. SECO describes how a problem statement is encoded into an interpreted structure according to the world semantics and the mathematical semantics, and how this structure can either be specified into an algorithm when congruent with one or recoded into a deep structure thanks to mathematical semantics in order to solve a semantically incongruent problem. We illustrated its ability to explain a wide range of effects by confronting SECO, post hoc, to previous studies presenting challenging results that had yet to be accounted for within a unified framework.

The idea that there exist different possible encodings of a situation described in a specific problem is central in the SECO model, yet this view appeared only recently in the literature. Ever since Riley et al.'s (1983) work and their taxonomy of additive word problems, the view that a word problem can be reduced to its objective mathematical structure and that two isomorphs of the same problem can thus be considered as equivalent in terms of difficulty for the solvers was abandoned in favor of an approach putting more emphasis on the way different arithmetic word problems are interpreted. It has for example been highlighted by Riley et al. that *combine* and *compare* problems can be approached very differently by the solvers, even when both are subtraction problems involving the same numerical values. However, in their view, each situation is attached to only one category in a taxonomy encompassing all problems, therefore suggesting that there is only one way to interpret a given situation. Similarly to Socrates' depiction of the human ability to "separate things according to their natural divisions, without breaking any of the parts the way a clumsy butcher does" (Plato, trans. 2009, p.64), this view presumes that there exists a natural breakdown of the situations depicted by the problems, and that each situation falls within an objective category. Within SECO, the interpretation of the problem statement varies depending on the solver's knowledge: a given situation may thus lead to different encodings. In order to solve an incongruent problem, a solver usually needs to recode the initial representation they have of it. The idea that an initial representation will be recoded to allow the use of a solving algorithm is one that was not covered by Bassok's semantic alignment framework. Bassok and colleagues' theory focuses on the abstraction of an interpreted structure during the initial encoding of a problem (Bassok, 2001), yet what happens when this initial encoding leads to failure hasn't been addressed, especially in cases in which a different representation of the situation could allow the solvers to find the solution. When the first interpreted structure cannot be specified into a valid solving algorithm, SECO covers the possibility that one recodes the situation and manages to solve the problem, in accordance with with empirical findings such as the ones reported in Gamo et al. (2010) or Thevenot and Oakhill (2005).

We propose to take a brief look at the empirical prospects opened by SECO. First, because it accounts for the part played by world semantics, SECO predicts that different individuals with different knowledge or experiences about the world

may tackle a problem differently. For instance, imagine if Hudson's (1983) problem about birds and worms had been framed in terms of smurfs and mushrooms ("There are 5 smurfs and 3 mushrooms, how many more mushrooms than smurfs are there?"). Children who are familiar with the Smurfs comic series will know that each smurf has his or her own mushroom to live in (there are no housemates in the Smurf village!). Thus, SECO predicts that these children may be more likely to find the solution to the problem, because their world semantics about smurfs and their individual mushrooms will help them to construct a paired encoding in which each house is assigned to one smurf (see Case study 5 for more details on why this should facilitate the solving process). More generally, SECO makes the prediction that cultural differences in the world semantics evoked by a given problem statement may influence participants' interpreted structure and the subsequent strategies they will use to solve a problem. For instance, it is believed that Indonesian and English speakers tend to represent durations as linear distances (e.g. a long time), whereas Spanish and Greek speakers tend to represent durations as definite quantities (e.g. *mucho tiempo*) (Casasanto et al., 2004). Thus, SECO predicts that English and Greek speakers may perform differently on the duration problems used by Gamo et al. (2010) and described in the 3rd case study.

Second, SECO predicts that modifying the world semantics evoked by a problem may influence the interpreted structure encoded. Such representational differences could be measured by asking participants to produce drawings of the problems they solved (e.g. Edens & Potter, 2008). Similarly, recognition tasks may provide a way to probe participants' representation of the problems (e.g. Hegarty, Mayer, & Monk, 1995; Mani & Johnson-Laird, 1982; Verschaffel, 1994), to assess whether their interpreted structures differed depending on the problem statements. Third, a central point in SECO is the recoding pathway, according to which one can recode an interpreted structure into a new representation at a certain cost. This cost can be measured by higher error rates on problems needing a recoding and higher response times on problems successfully recoded (Gros, Sander, & Thibaut, 2019). Future works might even assess the increase in cognitive load associated to this process by measuring physiological responses such as pupil dilation during the recoding of semantically incongruent problems. Fourth, the existence of the expert encoding pathway may be tested by presenting experts with different problem statements: SECO predicts that experts' performance on problems

requiring a recoding may decrease less than that of lay solvers, due to the possibility for experts to directly encode the problems' deep structure, even on incongruent problems . Fifth, SECO accounts for the fact that students may experience difficulty trying to solve a problem if they either lack the relevant world semantics, the relevant mathematical semantics, the ability to recode a semantically incongruent representation or the ability to compute the solving algorithm. Moreover, SECO predicts that different errors will be associated with these different shortcomings. By testing separately students' mathematical knowledge, their world knowledge about the entities described in the problem statement and their ability to compute specific algorithms, SECO can be used to pinpoint and address distinct sources of difficulties.

By providing a finer-grained depiction of solvers' reasoning, SECO can inform future works on the encoding, recoding and solving of arithmetic word problems. The conception of experiments testing the aforementioned predictions should help determine the explanatory power of SECO, either bolstering its claims or leading to the development of new alternative models.

Semantic congruence as an educational lever to tackle arduous notions

The current paper defines semantic congruence and suggests that difficulties might arise when the world semantics evoked by a problem statement is semantically incongruent with the problem's solving algorithm. In this view, semantic incongruence is a source of interferences and should be overcome by the learners to efficiently solve the encountered problems. Therefore, developing new methods to help students modulate the influence of world semantics in order to directly access the deep structure of the problems could be especially promising. Still, moderating the influence of world semantics is not trivial, since our knowledge about the world has been shown to be deeply involved in our reasoning, be it relevant or not (Bassok, 2001; Bassok et al., 1998; Gros, Sander, & Thibaut, 2016; Gros, Thibaut, & Sander, 2017).

However, world semantics can also have a facilitative influence. Depending on the semantics attached to a problem, solvers will access a congruent solving algorithm more easily than they would with another problem statement. It has been

shown that understanding the situation described in a problem statement can be enough to successfully solve a problem, even for children who did not receive any prior explicit instruction regarding the arithmetic notions required (Carpenter & Moser, 1982; De Corte & Verschaffel, 1987; Ibarra & Lindvall, 1982; Thevenot & Barrouillet, 2015). If the depicted situation is the one “doing the thinking” (Hofstadter & Sander, 2013, p. 432) then the effort is minimal. Depending on the semantics imbued in a situation, its representation might be more or less congruent with the deep structure of the problem and thus render it more or less easy to solve. In this regard, one can imagine that an abstruse mathematical theorem might seem almost self-evident if presented in the appropriate semantic setting.

Designing such situations aiming at fostering the understanding of a complex notion may be achieved through conceptual rewording, as suggested by Vicente et al. (2007). In their study, they highlighted that rewording problem statements in a way that makes more explicit the semantic relations between the problems’ entities, is beneficial to the solvers. Indeed, difficult problems (*i.e.*, problems that had “to be solved in a different than the actual sequence of the events denoted in the problem”, Vicente et al., 2007, p. 837) benefited from conceptual rewording, which referred to situations in which the underlying semantic relations between the given and unknown sets were made more explicit than in the standard version. On the other hand, situational rewording (*i.e.*, when a problem statement is presented in a more enriched and elaborated way, *e.g.*, causal relations between events made more explicit) led to no improvement compared to the standard version. In SECO’s view, conceptual rewording was beneficial because it highlighted the mathematical dependencies between quantities, and thus favored the mapping of the world semantics onto the relevant mathematical semantics. On the other hand, simply enriching the semantics of the situation had no effect on the mapping between the statement and mathematical representation. Thus, rewording will work when it aids in building a representation of the mathematical semantics that is congruent with the world semantics.

As a consequence, a crucial application of the SECO model resides in the development of educational interventions treating mathematical learning difficulties by resorting to world semantics in order to help understand and overcome some of the learners’ impairments regarding arithmetic understanding. Because SECO differentiates between world semantics, mathematical semantics

and algorithms, it can provide a detailed account of the potential difficulties encountered by students when learning to solve arithmetic word problems. The different components described in the model and the processes that link them are all potential candidates from which specific difficulties may stem. Using SECO, it is possible to differentiate between, for example, a lack of mathematical semantics (*e.g.* not knowing about the commutative property of multiplication) and difficulties in computing algorithms (*e.g.* not being able to calculate 3×50), in order to design targeted interventions which would help learners overcome their specific difficulties.

Gaining expertise

One of the distinctive features of the SECO model is that it provides an account of the part played by expertise in the solving of arithmetic word problems. The expert encoding pathway as introduced in SECO accounts for the idea, already developed by Chi et al. (1981), that solvers with sufficient expertise may be able to directly encode the deep structure of a problem, regardless of the world semantics it evokes. Data gathered regarding sorting and solving strategies depending on the learner's level of expertise, in line with Chi et al.'s (1981) seminal work, provide converging evidence regarding this view (*e.g.* Schoenfeld & Herrmann, 1982; Silver, 1981). Thus, a crucial educational issue is to promote learners' ability to reach a level of expertise allowing them to directly perceive a problem's deep structure, without first encoding an interpreted structure influenced by their everyday knowledge about the problem's entities.

However, since even expert solvers have been shown to sometimes rely on superficial features to determine their solving strategies (Blessing & Ross, 1996; Novick, 1988), experts' ability to ignore the influence of world semantics in all situations should not be taken for granted. In fact, recent evidence collected on problems similar to those described in our fourth case study suggests that general expertise in mathematics may not be sufficient to overcome the effects of semantic incongruence (Gros, Sander, & Thibaut, 2019). In this paper, it was shown that university-educated adults and expert mathematicians alike were more likely to deem an arithmetic word problem unsolvable when its solution was semantically incongruent with the world semantics evoked by the problem than when the two were semantically congruent. Does this mean that direct encoding of the deep

structure is unattainable? Not necessarily. It could be argued that the influence of world semantics is so pervasive that only specific expertise on the type of problem that is being solved (as compared with general expertise in mathematics) may provide the ability to directly encode the deep structure of the problem. In an educational perspective, the overall goal is to teach students either how to directly perceive the deep structure of the problems they encounter, or at least to efficiently recode an ineffective interpreted structure.

This raises the question of how one may develop such a level of expertise. Although conceptual rewording can be used to make a problem easier to solve, it does not necessarily mean that the solvers will learn to solve other problems which haven't been reworded. Correct answers are worth little if not associated to an increase in expertise. However, deliberately engaging in semantic recoding on multiple occasions on problems sharing the same deep structure may be a path to reach this goal. In Gamo et al.'s (2010) study, students' performance improved after they were explicitly told to compare "duration problems" and "number of elements problems", and taught how to semantically recode the number of elements problems to use the 1-step algorithm to solve them. As suggested by the rich literature on deliberate practice (Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005; Ericsson, 2004, 2008; Ericsson, Krampe, & Tesch-Römer, 1993; Lehtinen, Hannula-Sormunen, McMullen, & Gruber, 2017; Ward, Hodges, Starkes, & Williams, 2007) repeated training focused on specific tasks such as semantic recoding may be a promising path to develop top-level expertise. In this perspective, we know ever since Gick and Holyoak's work (1983) on analogical transfer that using different examples describing analogous situations can help represent their common structure (see also Braithwaite & Goldstone, 2015; Kotovsky & Gentner, 1996; Richland, Stigler, & Holyoak, 2012). It thus seems realistic to identify, for any type of problem, which problem statement as well as which sequence of training problems might be the most beneficial to help learners abstract a representation as close to the deep structure as possible. A congruence fading process akin to concreteness fading (Fyfe, McNeil, Son, & Goldstone, 2014) could thus help learners abstract the deep structure of the problems by resorting to increasingly incongruent examples. An interesting venue to capitalize on such effects would be to alternatively present problems attached to different world semantics congruent with different representations, in order to help learners switch

from an initial representation to another one, more efficient with regard to the resolution of the problem. Such scaled sequences of problems could be especially efficient if adapted to each learner through the use of Technology Enhanced Learning (Paquette, Léonard, Lundgren-Cayrol, Mihaila, & Gareau, 2006; Shute & Zapata-Rivera, 2012; Tchounikine, 2011). Although these propositions are only hypothetical at this stage, we consider these prospects to be promising leads for conducting further research and for helping foster transfer in mathematics education.

Broader application of the SECO model

An idea at the heart of the SECO model is that the congruence or the incongruence between the world knowledge elicited by a problem statement on one hand and the formal structure of the problem on the other hand can account for solvers' successes and failures, as well as for their need to recode their representations in incongruent situations. We believe this approach can also bear fruits if applied to other educational fields, such as mental arithmetic and non-mathematical problem solving.

Regarding arithmetic non-word problems, studies have shown that embedding an algorithm in a problem statement carrying world semantics may facilitate its computation (Baranes, Perry, & Stigler, 1989; Koedinger & Nathan, 2004; Stern & Lehrndorfer, 1992). SECO details how, depending on the congruence between world semantics and mathematical semantics, the solving process can be either favored or hindered by such an embedment. If an algorithm is embedded in a problem statement carrying congruent world semantics, then finding the solution should be easier. However, SECO also predicts that a problem statement carrying world semantics incongruent with the algorithm itself should have the opposite effect. Additionally, basic arithmetic operations carry a semantic meaning even when they are not framed within a problem statement (Bell, Swan, & Taylor, 1981; Fischbein, 1989; Fischbein, Deri, Nello, & Marino, 1985; Graeber, Tirosh, & Glover, 1989; Lakoff & Núñez, 2000; Tirosh & Graeber, 1991). According to Fischbein et al.'s (1985) view, arithmetic operations such as multiplication and division are attached to tacit models imposing constraints on their computation that have no mathematical relevance. For example, they argue that seeing division as the sharing of a collection of objects into a number of equal sub-collections implies that the

divisor must be a whole number and that the quotient must be smaller than the dividend. SECO addresses what happens when the world semantics evoked by the problem statement is incongruent with the objective mathematical structure of the problem. For example, believing that “to divide is to equally share” might lead solvers to rely on semantic knowledge regarding equitable sharing, making it harder to find the solution to arithmetic problems that go against this belief, such as $8 \div 0.5$. In this view, SECO can guide the analysis of the solvers’ activity when faced with such semantic incongruence by showing how the world semantics imbued in the operations themselves evoke an interpreted structure that is incompatible with the solving procedure.

By describing the influence of world semantics on arithmetic problem solving, SECO also underlines the facilitative role that a semantically congruent context may have on arithmetic reasoning in general. Interestingly, the influence of context on the understanding of arithmetic principles has been the focus of several works studying principles such as commutativity or inversion (see Prather & Alibali, 2009, for a review). According to Resnick’s (1992, 1994) theory of how mathematical competence is built, arithmetic understanding should emerge following a concrete-to-abstract transition, shifting from an initial object context to a verbal context, then a symbolic context, and then finally to an abstract context. For instance, learning about the commutative property of the addition of two sets of objects may not necessarily mean that learners will immediately be able to transfer this knowledge to the addition of numbers in general (Prather & Alibali, 2009). In a study about 7- to 9-year-olds’ understanding of arithmetic principles, Canobi (2005) showed that some children were helped by a concrete aid to display an understanding of a particular conceptual relation. She showed that some of the participants had an easier time explaining mathematical notions (subtraction complement and inversion principles) when presented with concrete objects instead of abstract numbers. Regarding the principle of commutativity, Cowan and Renton (1996) found that 6- to 9-year-olds showed a better understanding of commutativity in an object context or in a symbolic context, rather than in an abstract context. In other words, performance on mathematically identical tasks depended on the context in which the tasks were presented. Similarly to how SECO describes that the semantic embedding of a word problem can influence learners’ ability to find its solution, children’s performance in Cowan and Renton’s

study depended on the context of the task. Finally, in a study on arithmetic problem solving, Jordan, Huttenlocher, and Levine (1992) also found that disparities between middle-income children and low-income children disappeared when the questions were asked using objects rather than when the problems were only posed verbally.

Although few studies have been designed to specifically target the effects of context on principle understanding, and some have reported null effects (e.g. Canobi, Reeve, & Pattison, 2003), most works in the literature are compatible with the theory that children first learn the meaning of arithmetic principles in a grounded context before moving up to higher degrees of abstraction (Prather & Alibali, 2009). A parallel can be drawn with SECO, which accounts for the embedding of an arithmetic problem within a problem statement evoking specific world semantics. In both cases, solvers need to learn how to move away from a grounded encoding and towards a more abstract representation of the situation. We mentioned earlier how the use of increasingly semantically incongruent examples may complement a learning strategy based on concreteness fading (Fyfe et al., 2014), to guide learners from a concrete grasp of a problem to a more abstract understanding of its solution principle. It may be possible to develop a similar strategy in arithmetic learning, by progressively varying the semantic congruence between the concrete situations presented to the learners and the arithmetic notions to be taught.

Regarding problem solving in general, it is well established that the knowledge one has about the entities depicted in a problem can constrain their ability to find a solution (Clement & Richard, 1997; Duncker, 1945; Griggs & Cox, 1982; Kotovsky et al., 1985). Consider, for example, the physics problem consisting in asking whether when a car performs a circular motion at constant speed, its left-side door moves at the same speed as its right-side door or not. Most people trying to solve this problem will use their experience with cars and their world knowledge about rigid objects and represent the two doors of the car as parts of the same object. A common erroneous answer is that when a car moves, every part of the car moves at the same speed, since every passenger departs and arrives at the same time. We believe that the principles underlying SECO can help understand the solvers' reasoning on such a physics problem. In this case, the world semantics used to encode the problem into an interpreted structure will hide some physically

relevant aspects of the problem. Unless participants use physics semantics to perform a semantic recoding of the problem that dissociates the two doors as moving along two different circular paths which entails that they do not necessarily travel the same distance, their world semantics will lead them to the erroneous conclusion that the doors travel at the same speed. The notion that congruence between world knowledge and conceptual knowledge associated with a domain of instruction (*e.g.* mathematical semantics in the case of arithmetic problems, physics semantics for mechanics problems, and so on) can constrain the representation of situations and alter one's reasoning, unless a reinterpretation of the situation happens, seems to be a promising idea. In this regard, the scope of the SECO model could be extended in order to describe the encoding and recoding of situations from different domains of instruction, according to the world semantics and to the domain-related semantics influencing the solvers' interpretation.

Conclusion

The question of how one reasons when solving an arithmetic word problem is a major issue of mathematical education. Understanding the determinants of problem solving is a crucial step in order to identify the difficulties that should be addressed when teaching mathematics. The SECO model provides ground for a distinction between the mathematical semantics, the world semantics and the algorithms, as well as the way they interact and apply to familiar situations. Those interactions specify the steps involved in the encoding and the recoding of arithmetic word problems. Being able to foster a semantic recoding in order to improve analogical transfer would be a major step forward in the field of arithmetic teaching, and might help pupils overcome some of their numerous difficulties regarding word problem solving (Gamo et al., 2010; Hegarty, Mayer, & Green, 1992; Richland et al., 2012; Thevenot & Barrouillet, 2015; Verschaffel & De Corte, 1997). Strengthening our grasp of the effects of semantic congruence and incongruence could thus pave the way towards the development of new teaching strategies, building on world and mathematical semantics to guide the students towards a more abstract and more efficient understanding of the encountered problems, contributing to their conception of mathematical notions (Richland et al., 2012).

References

- Baranes, R., Perry, M., & Stigler, J. W. (1989). Activation of real-world knowledge in the solution of word problems. *Cognition and Instruction*, 6(4), 287-318. https://doi.org/10.1207/s1532690xci0604_1
- Bassok, M. (2001). Semantic alignments in mathematical word problems. *The analogical mind: Perspectives from cognitive science*, 401-433.
- Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology*, 35, 99-134. <https://doi.org/10.1006/cogp.1998.0675>
- Bassok, M., & Olseth, K. (1995). Object-based representations: Transfer between cases of continuous and discrete models of change. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1522-1538. <https://doi.org/10.1037/0278-7393.21.6.1522>
- Bassok, M., Pedigo, S. F., & Oskarsson, A. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 343-352.
- Bassok, M., Wu, L. L., & Olseth, K. L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory and Cognition*, 23, 354-367.
- Bell, A., Swan, M., & Taylor, G. (1981). Choice of operation in verbal problems with decimal numbers. *Educational studies in Mathematics*, 12(4), 399-420.
- Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3), 792. <https://doi.org/10.1037/0278-7393.22.3.792>
- Braithwaite, D. W., & Goldstone, R. L. (2015). Effects of variation and prior knowledge on abstract concept learning. *Cognition and Instruction*, 33(3), 226-256. <https://doi.org/10.1080/07370008.2015.1067215>
- Canobi, K. H. (2005). Children's profiles of addition and subtraction understanding. *Journal of Experimental Child Psychology*, 92(3), 220-246. <https://doi.org/10.1016/j.jecp.2005.06.001>
- Canobi, K. H., Reeve, R. A., & Pattison, P. E. (2003). Patterns of knowledge in children's addition. *Developmental Psychology*, 39(3), 521. <https://doi.org/10.1037/0012-1649.39.3.521>

- Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.
- Carpenter, T. P., & Moser, J. M. (1982). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective*, (pp. 9–24). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, I., ... Gil, D. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. In: K. Forbus, D. Gentner, T. Regier, (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, (pp. 575–580). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology*, *19*(2), 151-165. <https://doi.org/10.1002/acp.1106>
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.
- Clement, E., & Richard, J.-F. (1997). Knowledge of domain effects in problem representation: The case of Tower of Hanoi isomorphs. *Thinking & Reasoning*, *3*(2), 133-157. <https://doi.org/10.1080/135467897394392>
- Coquin-Viennot, D., & Moreau, S. (2003). Highlighting the role of the episodic situation model in the solving of arithmetical problems. *European Journal of Psychology of Education*, *3*, 267-279.
- Cowan, R., & Renton, M. (1996). Do they know what they are doing? Children's use of economical addition strategies and knowledge of commutativity. *Educational Psychology*, *16*(4), 407-420.
- Cummins, D. D. (1991). Children's Interpretations of Arithmetic Word Problems. *Cognition and Instruction*, *8*(3), 261-289. https://doi.org/10.1207/s1532690xci0803_2
- Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, *20*(4), 405-438. [https://doi.org/10.1016/0010-0285\(88\)90011-4](https://doi.org/10.1016/0010-0285(88)90011-4)
- Davidson, J. E., & Sternberg, R. J. (2003). *The psychology of problem solving*. New York: Cambridge University Press.

- Davis-Dorsey, J., Ross, S. M., & Morrison, G. R. (1991). The role of rewording and context personalization in the solving of mathematical word problems. *Journal of Educational Psychology*, 83(1), 61-68. <https://doi.org/10.1037/0022-0663.83.1.61>
- De Corte, E., & Verschaffel, L. (1987). The effect of semantic structure on first graders' strategies for solving addition and subtraction word problems. *Journal for research in mathematics education*, 363-381. <https://doi.org/10.2307/749085>
- 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. <https://doi.org/10.1037/0022-0663.77.4.460>
- Duncker, K. (1945). On problem-solving (L. S. Lees, Trans). *Psychological Monographs*, 58 (5, Whole No. 270).
- 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. <https://doi.org/10.1111/j.1949-8594.2008.tb17827.x>
- Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic medicine*, 79(10), S70-S81.
- Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Academic emergency medicine*, 15(11), 988-994. <https://doi.org/10.1111/j.1553-2712.2008.00227.x>
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3), 363. <https://doi:10.1037/0033-295X.100.3.363>
- Fischbein, E. (1989). Tacit models and mathematical reasoning. *For the learning of mathematics*, 9(2), 9-14.
- Fischbein, E., Deri, M., Nello, M. S., & Marino, M. S. (1985). The role of implicit models in solving verbal problems in multiplication and division. *Journal for research in mathematics education*, 3-17. <https://doi.org/10.2307/748969>

- Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, *26*(1), 9-25. <https://doi.org/10.1007/s10648-014-9249-3>
- Gamo, S., Sander, E., & Richard, J.-F. (2010). Transfer of strategy use by semantic recoding in arithmetic problem solving. *Learning and Instruction*, *20*, 400-410. <https://doi.org/10.1016/j.learninstruc.2009.04.001>
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford: Oxford University Press.
- Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, *59*(1), 47-59. <https://doi.org/10.2307/1130388>
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive psychology*, *15*(1), 1-38. [https://doi.org/10.1016/0010-0285\(83\)90002-6](https://doi.org/10.1016/0010-0285(83)90002-6)
- Goswami, U., & Brown, A. L. (1990). Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition*, *35*(1), 69-95. [https://doi.org/10.1016/0010-0277\(90\)90037-K](https://doi.org/10.1016/0010-0277(90)90037-K)
- Graeber, A. O., Tirosh, D., & Glover, R. (1989). Preservice teachers' misconceptions in solving verbal problems in multiplication and division. *Journal for research in mathematics education*, *20*(1), 95-102. <https://doi.org/10.2307/749100>
- Greer, B. (1992). Multiplication and division as models of solutions. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 276-295). New York: Macmillan.
- Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*(3), 407-420.
- Gros, H., Sander, E., & Thibaut, J. P. (2016). "This problem has no solution": When closing one of two doors results in failure to access any. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1271-1276). Austin, TX: Cognitive Science Society.
- 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*, online first. <https://doi.org/10.3758/s13423-019-01628-3>
- Gros, H., Thibaut, J.-P., & Sander, E. (2017). The nature of quantities influences the representation of arithmetic problems: Evidence from drawings and solving

- procedures in children and adults. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. London, UK: Cognitive Science Society.
- Große, C. S., & Renkl, A. (2006). Effects of multiple solution methods in mathematics learning. *Learning and Instruction*, 16(2), 122-138. <https://doi.org/10.1016/j.learninstruc.2006.02.001>
- Guthormsen, A. M., Fisher, K. J., Bassok, M., Osterhout, L., DeWolf, M., & Holyoak, K. J. (2016). Conceptual integration of arithmetic operations with real-world knowledge: Evidence from event-related potentials. *Cognitive science*, 40(3), 723-757. <https://doi.org/10.1111/cogs.12238>
- Gvozdic, K., & Sander, E. (2017). Solving additive word problems: Intuitive strategies make the difference. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. London, UK: Cognitive Science Society.
- Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology*, 84(1), 76.
- Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of educational psychology*, 87(1), 18. <https://doi.org/10.1037/0022-0663.87.1.18>
- Hofstadter, D. R., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. New York: Basic Books.
- Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. *Child Development*, Vol. 54, No. 1, 84-90. <https://doi.org/10.2307/1129864>
- Ibarra, C. G., & Lindvall, C. M. (1982). Factors associated with the ability of kindergarten children to solve simple arithmetic story problems. *The Journal of Educational Research*, 75(3), 149-156. <https://doi.org/10.1080/00220671.1982.10885372>
- Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive science*, 4(1), 71-115.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Harvard University Press.

- 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>
- Jordan, N. C., Huttenlocher, J., & Levine, S. C. (1992). Differential calculation abilities in young children from middle-and low-income families. *Developmental Psychology*, *28*(4), 644. <https://doi.org/10.1037/0012-1649.28.4.644>
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological review*, *92*(1), 109.
- Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science*, *32*(2), 366-397. <https://doi.org/10.1080/03640210701863933>
- Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The journal of the learning sciences*, *13*(2), 129-164. https://doi.org/10.1207/s15327809jls1302_1
- Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive psychology*, *17*(2), 248-294. [https://doi.org/10.1016/0010-0285\(85\)90009-X](https://doi.org/10.1016/0010-0285(85)90009-X)
- Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*(6), 2797-2822. <https://doi.org/10.1111/j.1467-8624.1996.tb01889.x>
- Kouba, V. L. (1989). Children's Solution Strategies for Equivalent Set Multiplication and Division Word Problems. *Journal for Research in Mathematics Education*, *20*(2), 147-158. <https://doi.org/10.2307/749279>
- Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: how the embodied mind brings mathematics into being* (Nachdr.). New York, NY: Basic Books.
- Lehtinen, E., Hannula-Sormunen, M., McMullen, J., & Gruber, H. (2017). Cultivating mathematical skills: From drill-and-practice to deliberate practice. *ZDM*, *49*(4), 625-636. <https://doi.org/10.1007/s11858-017-0856-6>
- Leikin, R., & Lev, M. (2007). Multiple solution tasks as a magnifying glass for observation of mathematical creativity. In *Proceedings of the 31st international conference for the psychology of mathematics education*, *3*,

- 161–168. The Korea Society of Educational Studies in Mathematics Seoul, Korea.
- Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition*, *10*(2), 181-187. <https://doi.org/10.3758/BF03209220>
- Nesher, P., Greeno, J. G., & Riley, M. S. (1982). The development of semantic categories for addition and subtraction. *Educational Studies in Mathematics*, *13*(4), 373–394.
- Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-Hall.
- Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, memory, and cognition*, *14*(3), 510.
- Paquette, G., Léonard, M., Lundgren-Cayrol, K., Mihaila, S., & Gareau, D. (2006). Learning design based on graphical knowledge-modelling. *Journal of Educational Technology & Society*, *9*(1).
- Plato (2009). *Plato's Phædrus*. Millis, Mass.: Agora Publications.
- Prather, R. W., & Alibali, M. W. (2009). The development of arithmetic principle knowledge: How do we know what learners know?. *Developmental Review*, *29*(4), 221-248. <https://doi.org/10.1016/j.dr.2009.09.001>
- Resnick, L. B. (1992). From protoquantities to operators: Building mathematical competence on a foundation of everyday knowledge. *Analysis of arithmetic for mathematics teaching*, *19*, 275-323.
- Resnick, L. B. (1994). Situated rationalism: Biological and social preparation for learning. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 474–493). Cambridge, UK: Cambridge University Press.
- Reusser, K. (1988). Problem solving beyond the logic of things: Contextual effects on understanding and solving word problems. *Instructional Science*, *17*(4), 309-338.
- Reusser, K. (1989). *Textual and situational factors in solving mathematical word problems* (Research Rep. No. 7). Bern, Switzerland: University of Bern, Department of Educational Psychology.
- 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*, Vol. II. New York: Pergamon Press.
- Reusser, K. (1993). Tutoring systems and pedagogical theory: Representational tools for understanding, planning, and reflection in problem solving. *Computers as cognitive tools, 1*, 143-177.
- Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the conceptual structure of mathematics. *Educational Psychologist, 47*(3), 189-203. <https://doi.org/10.1080/00461520.2012.667065>
- Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. *The development of mathematical thinking*, 153-196.
- Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(4), 629.
- Ross, B. H., & Bradshaw, G. L. (1994). Encoding effects of reminders. *Memory & Cognition, 22*(5), 591-605.
- Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Erlbaum
- Sander, E., & Richard, J.-F. (2005). Analogy and transfer: Encoding the problem at the right level of abstraction. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, Stresa, Italy, pp. 1925-1930.
- Schank, R. C. (1975). The role of memory in language processing. In: C. N. Cofer (ed), *The Structure of Human Memory* (pp. 162-189). San Francisco.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans and understanding: An inquiry into human knowledge structures*. Oxford, UK: Lawrence Erlbaum.
- Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(5), 484. <https://doi.org/10.1037/0278-7393.8.5.484>
- Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. Durlach (Ed.), *Adaptive technologies for training and education* (pp. 7-27). New York, NY: Cambridge University Press.

- Silver, E. A. (1981). Recall of mathematical problem information: Solving related problems. *Journal for research in mathematics education*, 54-64. <https://doi.org/10.3758/10.2307/748658>
- Squire, S., & Bryant, P. (2002). From sharing to dividing: Young children's understanding of division. *Developmental Science*, 5(4), 452-466. <https://doi.org/10.1111/1467-7687.00240>
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum.
- Staub, F. C., & Reusser, K. (1992). The role of presentational factors in understanding and solving mathematical word problems. In *meeting of the American Educational Research Association*, San Francisco.
- Staub, F. C., & Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension* (pp. 285-305). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stern, E., & Lehrndorfer, A. (1992). The role of situational context in solving word problems. *Cognitive Development*, 7(2), 259-268.
- Tchounikine, P. (2011). Educational software engineering. In *Computer Science and Educational Software Design* (pp. 111-122). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20003-8_6
- Thevenot, C. (2010). Arithmetic word problem solving: Evidence for the construction of a mental model. *Acta Psychologica*, 133(1), 90-95. <https://doi.org/10.1016/j.actpsy.2009.10.004>
- Thevenot, C. (2017). Arithmetic Word Problem Solving: The Role of Prior Knowledge. In D. C. Geary, D. B. Berch, R. J. Ochsendorf, & K. M. Koepke (Éd.), *Acquisition of Complex Arithmetic Skills and Higher-Order Mathematics Concepts* (pp. 47-66). <https://doi.org/10.1016/B978-0-12-805086-6.00003-5>
- Thevenot, C., & Barrouillet, P. (2015). Arithmetic word problem solving and mental representations. *The Oxford Handbook of Numerical Cognition*, 158.
- Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representations in arithmetic word problem solving. *Quarterly Journal of Experimental Psychology-A*, 58(7), 1311-1323. <https://doi.org/10.1080/02724980443000593>

- 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. <https://doi.org/10.1080/09541440500412270>
- Tirosh, D., & Graeber, A. O. (1991). The effect of problem type and common misconceptions on preservice elementary teachers' thinking about division. *School Science and Mathematics, 91*(4), 157-163.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Vergnaud, G. (1983). Multiplicative structures. *Acquisition of mathematics concepts and processes, 127-174*.
- Verschaffel, L. (1994). Using retelling data to study elementary school children's representations and solutions of compare problems. *Journal for Research in Mathematics Education, 141-165*. <https://doi.org/10.2307/749506>
- Verschaffel, L., & De Corte, E. (1997). Teaching realistic mathematical modeling in the elementary school: A teaching experiment with fifth graders. *Journal for Research in mathematics education, 577-601*.
- Vicente, S., Orrantia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology, 77*(4), 829-848.
- Ward, P., Hodges, N. J., Starkes, J. L., & Williams, M. A. (2007). The road to excellence: Deliberate practice and the development of expertise. *High ability studies, 18*(2), 119-153. <https://doi.org/10.1080/13598130701709715>

CHAPTER 2 – *WHAT WE COUNT DICTATES
HOW WE COUNT:
A TALE OF TWO ENCODINGS*



Presentation

In line with the idea, raised in Chapter 1, that the specific influence of world semantics needs to be taken into account when apprehending arithmetic word problem solving, the next chapter introduces the impact that our knowledge about the world has on our perception of cardinality and ordinality in numerical situations. The semantic distinction between cardinal and ordinal situations is a core aspect of this thesis, and it is the main focus of the experiments we carried out. Chapter 2 presents our theoretical motivations for exploring this dimension, and it defines what we call cardinal and ordinal quantities. The isomorphic problems we created to tackle this issue are described, as well as our predictions regarding the way they will be encoded, recoded and solved by participants.

A series of 6 experiments were designed to provide converging evidence for the decisive influence of cardinal and ordinal quantities on the interpretation of numerical situations by adult participants. The first experiment is a sorting task, in which participants are instructed to sort the problems into as many categories as they deem necessary. We predict that the classification pattern will reflect the semantic distinction we introduced between cardinal and ordinal problems. The second experiment uses a comparison paradigm in which participants are instructed to choose among a list of problems which ones can be solved according to the same principle as a source problem. We predict that the perception of the isomorphism between problems will depend on whether the quantities they use foster the same type of semantic encoding. Experiment 3 assesses the replicability of Experiment 2 in a paired comparison setting. We predict that instructing participants to directly compare two problems will not be sufficient to overcome the encoding difference between cardinal and ordinal problems. Experiment 4 regards the algorithms used to solve the problems. We introduce a new type of “hybrid” problems, meant to foster an ordinal encoding of problems featuring cardinal quantities. We predict that the choice of the solving algorithm will be guided by our manipulation of the problem statements’ semantic properties along the cardinal-ordinal dimension. Experiment 5 introduces a new type of problem alteration, meant to make it impossible to solve a problem encoded following a cardinal representation. A solvability assessment task is presented to the participants, in which they have to decide which problems can be solved and which problems have no solution. We predict that solvable problems

using cardinal quantities will be more often deemed unsolvable than problems using ordinal quantities. Experiment 6 is a conceptual replication of Experiment 5 with stronger constraints: the solution to the problems is directly given to the participants. We predict that providing the solution will not be enough to help participants overcome the difficulty arising from semantic incongruence.

Overall, these 6 experiments are meant to establish the relevance of the distinction between quantities emphasizing the cardinality of numbers and quantities emphasizing their ordinality. By confirming the validity of this distinction in Experiments 1, 2, and 3, by showing its influence on problem solving in Experiment 4, and by illustrating how it can lead to important mistakes in Experiments 5 and 6, we strove to bring substantial proof of its importance in human reasoning about numerical situations. The following chapters all build upon this work to investigate the extent to which cardinality and ordinality interfere with mathematical reasoning.

Experiments 5 and 6 have been reported in an article in the Proceedings of the Annual Meeting of the Cognitive Science Society (2016). This chapter as a whole is currently being reviewed for publication by the *Journal of Experimental Psychology: General*.

Abstract

We argue that *what* we count has a crucial impact on *how* we count, to the extent that even adults may have difficulty using elementary mathematical notions in concrete situations. Specifically, we investigate how the use of specific quantities (durations, heights, number of floors) may emphasize the ordinality of the numbers featured in a problem, whereas other quantities (collections, weights, prices) may emphasize the cardinality of the depicted numerical situations. We suggest that this distinction leads to the construction of one of two possible encodings, either a cardinal or an ordinal representation. This difference should in turn constrain the way we approach problems, influencing our mathematical reasoning in multiple activities. This hypothesis is tested in six experiments ($N = 916$), using different versions of multiple-strategy arithmetic word problems. We predict that the distinction between cardinal and ordinal quantities influences problem sorting (Experiment 1), perception of similarity between problems (Experiment 2), direct problem comparison (Experiment 3), choice of a solving algorithm (Experiment 4), problem solvability estimation (Experiment 5) and solution validity assessment (Experiment 6). The results provide converging clues shedding light into the fundamental importance of the cardinal versus ordinal distinction on adults' reasoning about numerical situations. Overall, we report multiple evidence that general, non-mathematical knowledge associated with the use of different quantities shapes adults' encoding, recoding and solving of mathematical word problems. The implications regarding mathematical cognition and theories of arithmetic problem solving are discussed.

Keywords

mathematical cognition · mental model · problem solving · representation · encoding; strategy choice · semantics.

Introduction

Mathematical notions are frequently used in everyday life (Northcote & Marshall, 2016). Be it to count the coins to give the bus driver, to calculate how long your ride will take, to decide whether to go up the stairs or take the elevator, to assess whether the added weight of everyone in the elevator is below the weight limit, to count how many cookies you need to buy for your coworkers, or to evaluate at which height you need to put the cookie jar so that the kids won't reach it. However, being able to perform those quick calculations on a daily basis requires the ability to use mathematical notions in very tangible situations (Schoenfeld, 2009). Mastering such a skill is the focus of a long – and sometimes painful – learning process taking place in school, notably involving mathematical word problems (Daroczy, Wolska, Meurers, & Nuerk, 2015; Stacey, 2005). But while using arithmetic notions in daily-life might seem effortless to most adults, we argue that some underlying pervasive influence remains: in fact, even when using the most elementary mathematical notions, we have reasons to believe that *what* we count has a deep impact on *how* we count. This entails that even adults may be significantly better at using certain fundamental mathematical notions in situations involving bus rides, elevator trips or cookie jar heights than in situations pertaining to bus ticket prices, elevator weight limits or cookie counting. In this paper, we argue that such concrete situations evoke one of two distinct conceptions of numerical situations; they emphasize either the *cardinal* property of numbers or their *ordinal* property. This difference leads to one of two possible encodings, which in turn constrain the way we represent word problems, influencing our mathematical reasoning in multiple activities.

An ontological distinction between cardinal and ordinal situations

This distinction between cardinality and ordinality is fundamental in mathematics, especially in set theory (Dantzig, 1945; Frege, 1980; Russell, 1919; Simon, 1997; Suppes, 1972). In common usage, ordinal numbers describe the numerical position of an object in an ordered sequence (i.e. 1st, 2nd, 3rd, etc.), whereas cardinal numbers refer to the general concept of quantity by designating the total number of entities within a set (Fuson, 1988; Wasner, Moeller, Fischer, & Nuerk, 2015). But this distinction has far-reaching implications beyond set theory and mathematics, and we believe that it influences the very way humans comprehend numerosity in the world.

According to Piaget, the relationship between cardinality and ordinality is

central to the notion of number (Fuson, 1988). Indeed, the two ideas are necessarily intertwined, since the cardinal and ordinal meanings of numbers are two sides of the same coin, two properties intrinsically tied to our use of numbers. As Fuson (1988, p. 363) puts it, “any ordinal number refers to the particular entity within a linear ordering that is preceded by a cardinal number one less than it”. However, this interdependency between ordinality and cardinality does not preclude that an ontological difference between conceiving of numbers as cardinal count values or as order labels could exist.

From a developmental perspective, the seminal work of Gelman and Gallistel (1986) on counting principles introduced the idea that the understanding of the cardinal and ordinal properties of numbers could develop separately in children. In their work, Gelman and Gallistel notably argued that children needed to learn the “stable-order-principle” and the “cardinal principle” in order to become proficient counters. Mastery of the stable-order-principle means that a child has learnt that the list of words used to count must be used in a definite and repeatable order; it can be linked to the development of the notion of ordinality. The cardinal-principle, on the other hand, refers to the understanding that the number name allocated to the final entity in a collection corresponds to the total number of entities being counted. Ever since this work, investigations have been conducted to study how children learn to master both properties of numbers.

Children’s ability to comprehend and use the notion of cardinality in counting was thus scrutinized by numerous studies who showed its slow and sequential development over the first years of life (e.g. Bermejo, 1996; Condry & Spelke, 2008; Le Corre & Carey, 2007; Sarnecka & Lee, 2009; Wynn, 1992). As for the development of the ordinal meaning of numbers, it has been suggested that children do not learn to use ordinal labels such as “first” and “second” before they are 4 or 5 years old (Fischer & Beckey, 1990; Miller, Major, Shu, & Zhang, 2000; Miller, Marcovitch, Boseovski, & Lewkowicz, 2015). In recent years, the comparison of the developmental trajectories of the sense of cardinality and the sense of ordinality has received increasing research interest, with converging results showing that children tend to use cardinality before succeeding in ordinality tasks (Colomé & Noël, 2012; Meyer, Barbiers, & Weerman, 2016; Wasner, Moeller, Fischer, & Nuerk, 2015). Further evidence for the importance of this distinction can be found in the brain; Delazer and Butterworth (1997) reported that a patient who suffered a cerebral lesion was left with

an impaired access to the cardinal meaning of numbers, but a selectively preserved access to their ordinal meaning. This dissociation between the two meanings of numbers was supported by EEG studies who found that event-related potentials differed both in terms of timing and topography between order and quantity tasks (Rubinsten, Dana, Lavro, & Berger, 2013; Turconi, Jemel, Rossion, & Seron, 2004). Overall, these studies show the importance of this distinction in human cognition. However, little is known about its influence on adult reasoning, as the distinction between cardinality and ordinality is rarely mentioned once counting procedures are considered to be acquired. This paper aims at filling this gap, by focusing on adults' ability to use basic mathematical notions in situations evoking either the cardinal or the ordinal aspect of numbers. Indeed, we believe that the presence of cardinal versus ordinal quantities in the wording of otherwise mathematically identical problems may lead even adults to build one of two different encodings of the depicted situations. Our claim is that this distinction will, in turn, lead to clear-cut differences in the way these problems are categorized, compared and solved. It builds upon the literature on arithmetic word problems, with a special focus on the contrasting frameworks developed to account for wording effects in mathematical reasoning.

Interpretative effects at play in arithmetic problem solving

We have known ever since Riley, Greeno and Heller's work (1983) on additive word problems that applying the same mathematical notions to different situations can present its own challenges. Indeed, a vast literature has shown that slight modifications in the wording of otherwise structurally identical mathematical word problems could result in significant performance disparities (Carpenter & Moser, 1982; Coquin-Viennot & Moreau, 2003; Cummins, Kintsch, Reusser, & Weimer, 1988; De Corte, Verschaffel, & De Win, 1985; Hudson, 1983; Greer, 1992; Nesher, Greeno, & Riley, 1982; Squire & Bryant, 2002; Thevenot & Oakhill, 2005; Vergnaud, 1983). However, the identification of the underlying reasoning processes accounting for such disparities remains a debated issue to this day.

One possible explanation for these disparities comes from the schema theory (Kintsch, 1988; Kintsch & Greeno, 1985; Rumelhart, 1980; Schank, 1975; Schank & Abelson, 1977), which proposes that our ability to identify the algorithmic solution of a word problem depends on our capacity to activate the appropriate schema in long-term memory. According to Kintsch and Greeno (1985), schemas are defined as propositional data structures that can be implemented with the numerical values of

any problem statement sharing the same structure. Through repeated exposure to problems sharing the same structure, those schemas are created and stored in long-term memory; they can then be activated and act as operator structures using the problem's values to deduce the problem's solving algorithm (Kintsch & Greeno, 1985). For instance, according to Riley et al.'s (1983) typology, any comparison problem involving a sentence such as "Tom has 3 more marbles than Joe", triggers the retrieval of a have-more-than propositional structure that can be implemented with the 3 numerical values corresponding to Joe and Tom's sets and to the difference between the two.

However, a number of shortcomings have been shown to hinder the schema theory's explanatory power of the interpretative effects that seem to be central in mathematical word problem solving. For example, Thevenot (2010) asked participants to solve arithmetic word problems and then presented them with an unexpected recognition task involving either the original problems, inconsistent problems that had never been solved or paraphrastic problems which respected the relational structure of the original problems but not their propositional structure. Participants erroneously recognized the paraphrastic problems more often than the inconsistent problems, despite the paraphrastic problems being less similar to the original problems in terms of propositional structure, which was at odds with what the schema theory would have predicted.

Other effects put the schema theory's explanation to the test. The benefit of placing the question of an arithmetic word problem at the beginning of a problem statement instead of at the end is a thoroughly demonstrated effect (Devidal, Fayol, & Barrouillet, 1997; Fayol, Abdi, & Gombert, 1987). According to the schema theory, it is due to the fact that the question helps activate the appropriate schema at the beginning of the process and thus facilitates calculations (Devidal et al., 1997). However, Thevenot, Devidal, Barrouillet and Fayol (2007) showed that placing the question prior to the text of a problem benefited more to children with poor mathematical skills than it did to children with a higher mathematical proficiency. This effect is challenging for the schema theory. Indeed, according to this theory, participants with poor mathematical skills possess fewer schemas in their long-term memory, and so placing the question at the beginning of a problem statement should help participants select the appropriate schema only if it belongs to their repertoire. Thus, this result contributed to undermine the relevance of the schema theory in

mathematical problem solving, by highlighting the limits of its predictions regarding the difference between students with high and low proficiency in mathematics.

However, ever since 1990, an alternative approach has emerged from Reusser's critical observations on the schema theory. According to Reusser, the schema model postulates a one-step mathematization process neglecting how the understanding of the situation described in the problem statement itself may influence the solving process. Consequently, he proposed the Situation-Problem-Solver, a model introducing the idea that the text of the problem statement is first translated into a situation model of the situation, before being "mathematized" into a problem model and finally translated into a solving algorithm (Reusser, 1990; Staub & Reusser, 1995). Reusser's use of the concept of "mental model", which was introduced by Johnson-Laird (1980; 1983) in the domain of reasoning and text comprehension, accounts for the creation of a non-mathematical representation, akin to the situation model described by Van Dijk and Kintsch (1983). This representation describes the entities involved in the problem statement and the relational structure they entertain (Staub & Reusser, 1995).

Yet, it can be argued that by relying on the notion of situation model, the Situation-Problem-Solver approach neglects some interpretative effects at play in mathematical word problem solving. Indeed, according to Johnson-Laird (2010), the structure of a situation model "corresponds to the structure of what it represents" (p.18244). In the Situation-Problem-Solver model, this translates into the notion that the text of a problem depicts one specific situation, thus leaving little to no room for the idea that alternative representations of a same problem may exist. In other words, this approach does not directly include the idea that different models of a same problem can be built, depending on the solver's viewpoint. Thus, neither the influence of the solvers' prior knowledge nor the possibility to go from one initial representation to a new encoding of a situation fall within the scope of the Situation-Problem-Solver approach. Posterior works have focused on these aspects and what they entail for theories of mathematical word problem solving.

Semantic induction in question

Bassok, Wu and Olseth (1995) showed that semantic knowledge linked to the different entities described in a problem statement could influence analogical transfer between problems. They taught participants the algorithmic solution of a problem whose cover story depicted either objects assigned to people (e.g. computers given

to secretaries) or people assigned to other people (e.g. doctors from one hospital assigned to doctors from another hospital). They then evaluated how well participants performed on problems sharing the same solution principle. Depending on whether the entities in the transfer problem had typically symmetrical roles (people and people) or typically asymmetrical roles (objects and people), participants' performance varied. Those results suggested that variations of semantic knowledge lead to different representations being abstracted. These "interpreted structures" either facilitated or hindered transfer depending on how well they mapped onto the problems' mathematical structure. Similarly, Bassok, Chase and Martin (1998) asked participants to create addition or division word problems using specific sets of entities, either linked by a functional semantic relation (e.g. the container/content relation between fruit baskets and oranges) or showcasing a collateral relation (e.g. oranges and apples belonging to the same superordinate "fruit" category). They showed that participants tended to propose division problems when the semantic knowledge induced by the entities evoked a functional relation, whereas they created addition problems when the entities were different kinds of fruits, or other collateral elements.

Thus, the semantic knowledge induced by the entities described in the problem statement seems to influence the representation that is encoded by the participants. Bassok (2001) theorized this process by stating that an interpreted mathematical structure is abstracted by the participants, based on the semantic relations depicted in the problem statements (e.g. container/content or assigned/receiver relations). This interpreted structure can be semantically aligned or misaligned with the objective mathematical structure of the problem, depending on whether the two structures can be mapped onto each other (Bassok, 2001). For instance, the interpreted structure of a problem involving oranges and baskets will be semantically aligned with division, whereas the interpreted structure of a problem involving oranges and apples will not. The variations in semantic alignment thus lead to performance differences that are not accounted for by the schema framework nor by the mental model theory. In support of this view, Bassok, Pedigo and Oskarsson's work (2008) on the priming of addition facts by different pairs of words is enlightening. They showed that categorically related words (e.g. tigers and cheetahs) prime addition facts, whereas unrelated words (e.g. lungs and statues), as well as functionally related words (e.g. bears and claws), do not exert such a priming. In other words, pairs of words semantically aligned with addition elicit an automatic

retrieval of addition facts, whereas words misaligned with the structure of addition do not. Taken together, those studies suggest that the semantics induced by the objects manipulated in the problem statements influence the representational processes at play.

Cardinal and ordinal representations in arithmetic problems

We argue that the semantic knowledge about the entities described in a problem can evoke either an ordinal or a cardinal representation of the situation described, subsequently shaping mathematical reasoning. Indeed, preliminary work on the perception of cardinal and ordinal situations by individuals who have already mastered the counting procedures was undertaken by Gamo et al. (2010) in a study with 4th and 5th graders. The authors created multiple-solution word problems sharing the same mathematical structure but differing in the type of quantities used in their problem statements. Their problems revolved around, respectively, the number of family members in a hotel (family problems), the price of a series of items (price problems) or the age of a protagonist (age problems) (see Table 1).

Table 1 Family, price and age problems used in Gamo, Sander, & Richard, 2010.

Family problem	Price problem	Age problem
In the Richards' family, there are 5 persons. When the Richards go on vacation with the Roberts, there are 9 persons at the hotel. In the Dumas' family, there are 3 people less than in the Richards' family. The Roberts go on vacation with the Dumas. How many will they be at the hotel?	John bought an 8-Euro exercise book, and scissors. He paid 14 Euros. A pen costs 3 Euros less than the exercise book. Paul bought scissors and a pen. How much did he pay?	Antoine attended painting classes at the art school for 8 years and stopped when he was 17 years old. Jean began at the same age as Antoine and attended the course for 2 years less. At what age did Jean stop attending the classes?

Gamo et al. hypothesized that family problems and price problems would both lead to a cardinal encoding of the situation described, whereas age problems would lead to an ordinal encoding of the situation. Consider the family problem (Table 1, left column). Gamo et al. showed that most participants solved it using a 3-step algorithm: $9 - 5 = 4$; $5 - 3 = 2$; $4 + 2 = 6$. Note, however, that this problem can also be solved with a one-step algorithm: $9 - 3 = 6$. But using this algorithm requires for the participant to realize that since the Roberts are present at the hotel during both vacations, then the difference between the number of people in the Richards' and in

the Dumas' families is equal to the difference between the total number of people at the hotel. Thus, calculating the number of people in the Roberts' family ($9 - 5 = 4$) or the number of people in the Dumas' family ($5 - 3 = 2$) is not necessary to find the solution.

Similarly, most participants use a 3-step algorithm to solve the price problem (Table 1, middle column). This algorithm consists in calculating the price of the pen and the price of the scissors, and then adding them up: $14 - 8 = 6$; $8 - 3 = 5$; $6 + 5 = 11$. But a 1-step algorithm can also be used to solve this problem: $14 - 3 = 11$. However, only participants who notice that John and Paul both bought scissors may also understand that the difference between the price of the exercise book and the price of the pen is equal to the difference between what John and Paul paid in total, and thus be able to find this 1-step algorithm. Lastly, the age problem (Table 1, right column) could also be solved using a 3-step algorithm ($17 - 8 = 9$; $8 - 2 = 6$; $9 + 6 = 15$) or a 1-step algorithm ($17 - 2 = 15$). However, Gamo et al. (2010) found that participants used the 1-step algorithm more often on such problems, even though all three problems were isomorphic and shared the exact same mathematical structure (see Fig. 1).

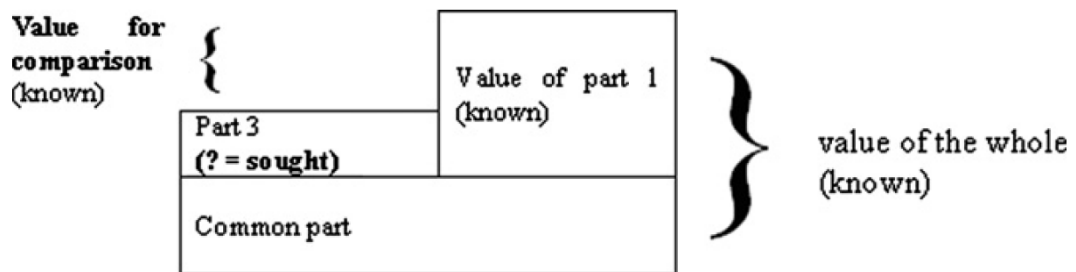


Fig. 1 The mathematical structure of the problems. Reprinted from “,” by S. Gamo, E. Sander, J.-F. Richard, 2010, *Learning & Instruction*, 40(5), p. 405.

The rationale behind this difference in participants' ability to use the shortest 1-step algorithm on all three problems was that the quantities used (family members, price, age) evoked different aspects of world knowledge that emphasized either the *cardinal* or the *ordinal* nature of the problems' values. The interpretation was that price or family problems tend to be encoded as sets of unordered, disconnected elements that can be grouped and whose values can be added with no ontological order. Such a cardinal encoding would foster the use of the 3-step algorithm consisting in calculating the individual value of each part making the whole, and then adding them

up. On the other hand, in age situations the age values are inherently ordered and can be represented as states (positions on an axis) and transitions (intervals on a timeline). This type of ordinal encoding would make it easier to compare two sets sharing a common part: if Jean and Antoine started attending painting classes at the same age, and Antoine attended the class for 2 years less than Jean, then making the inference that Jean was 2 years younger than Antoine at the end of their respective classes seems especially natural.

Current study

Our ambition for this paper is twofold. First, we aim at showing the critical influence that the distinction between cardinality and ordinality has on human understanding of situations involving numbers. We intend to show that, more than a mere developmental or mathematical question, this is an ontological issue regarding the way numerosity is perceived. Second, we intend to build on this semantic distinction to investigate adults' representations in the course of mathematical reasoning. Namely, we aim at showing that the nature of the entities staged in word problems interferes with their classification, comparison, encoding, recoding and solving, thus presenting a range of effects falling beyond the scope of the current models of arithmetic problem solving. We designed a series of 6 experiments scrutinizing how the cardinal versus ordinal aspects of numerical situations influence adults' reasoning in a variety of tasks.

Overview of the experiments

We believe that the selection of a specific type of quantity to create a problem statement has an influence on the encoding of the problem into either a cardinal or an ordinal representation, which in turn fosters the use of one of the two existing solving algorithms (see Fig. 2 for a graphical summary of this hypothesis).

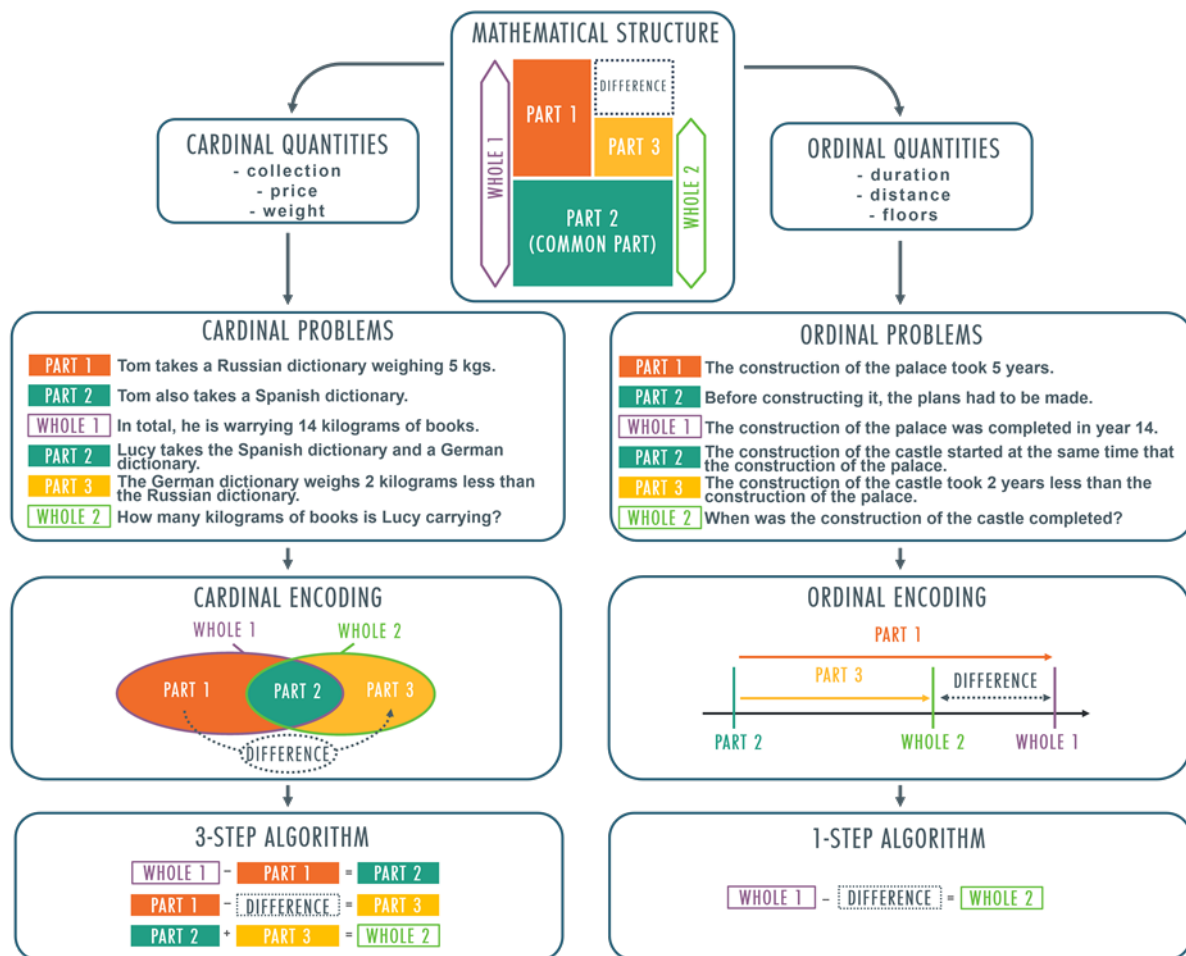


Fig. 2 Graphical summary of the content effects under scrutiny in this paper. Problems sharing the same mathematical structure are implemented with different quantities evoking different encodings, which in turn foster the use of one of the two solving algorithms.

We designed six experiments in order to examine the importance that the cardinal versus ordinal distinction holds in the representation of numerical situations. They were conducted with adults in order to underline the strength and pervasiveness of the reported effects. First, we devised a free sorting experiment to show the fundamental influence of the ordinal versus cardinal distinction on participants' spontaneous categorization of problems, while assessing the validity of our choice of materials. Second, we used an analogy identification experiment to determine whether the hypothesized semantic difference between cardinal and ordinal problems would predict how participants perceive the isomorphism between different problem statements. Third, we used a direct comparison task to evaluate the robustness of the second experiment's findings regarding the perception of the analogy between cardinal and ordinal problems. Fourth, we proposed a solving task with cardinal, ordinal and "hybrid" problems (cardinal problems presented in an ordinal context) to show that participants' choice of solving strategies depends on the semantics

introduced in the problems. Fifth, we used a solvability-assessment task to evaluate whether participants could solve problems whose unique solution was incompatible with their initial encoding of the problems. Sixth, we used a solution-validity-judgement task to assess whether participants' difficulty to find a problem's unique solution in the fifth experiment could be overcome by the presentation of a candidate solution to evaluate. All six experiments received the approval of the ethics committee of the University of Geneva. Altogether, these experiments intended to validate the distinction between cardinal and ordinal situations, to show its influence on the encoding of numerical situations even among adults well past their schooling years, and to evaluate whether its influence could be so robust as to interfere with participants' ability to use relatively basic arithmetic notions to solve 1-step arithmetic problems.

In an attempt to maximize the generality of our findings, we selected 3 types of quantities thought to evoke an ordinal encoding of the situation and 3 types of quantities that we assumed evoked a cardinal encoding. The rationale behind the selection of those specific quantities is presented below; the relevance of this classification will notably be assessed by our first experiment.

Overview of the selected ordinal quantities

Durations. In English, spatial metaphors expressing durations along a unidimensional space such as “a long time” are prevalent (Casasanto, 2008). The same is true for French, in which the word to designate an extensive period of time corresponds to the literal concatenation of the terms standing for “long” (“long”) and “time” (“temps”): “longtemps”. In western culture, including in France, the idea that time evolves along an axis or a timeline is deeply rooted in our understanding of the world (Bonato, Zorzi, & Umiltà, 2012; Boroditsky, 2011; Droit-Volet & Coull, 2015; Weger & Pratt, 2008). Despite cultural differences regarding the direction given to the axis of time (e.g. Fuhrman & Boroditsky, 2010), the fact remains that time is usually conceived of as unidirectional axis on which values are ontologically ordered. Contexts in which durations are mentioned or compared thus tend to increase the saliency of the ordinal meaning of numbers.

Heights. Height is a spatial, unidimensional concept. Due to our living in a world where gravity constantly exerts its influence on physical objects, we quickly learn that objects fall in a straight line and we tend to think of height as being an oriented, vertical axis, with a bottom and a top (e.g. Hood, 1998; Hood, Santos, & Fieselman,

2008; Kim & Spelke, 1992). Entities' heights can easily be compared or stacked alongside this axis, and there is an ontological order in the values representing the heights of different entities placed atop each other. Thus, mentions of height in a numerical situation tend to emphasize the ordinal meaning of the numbers involved.

Floors. Building floors are stacked on top of each other and they are ordered in a fashion that cannot be altered. If one is to imagine an elevator going from one floor to another, it immediately comes to mind that to go from the 1st floor to the 3rd floor, the elevator must pass the 2nd floor first (Clement & Richard, 1997). Thus, floors have an ontological order that we believe highlights the ordinality of the values used to count them.

Overview of the selected cardinal quantities

Collections. Collections refer to groups of disconnected elements that can be counted as parts of a set. Be it a collection of blue marbles in a bag, a set of iguanas in a terrarium or a group of pupils in a bus, collections of similar elements usually have no ontological order. Fittingly, most studies on the development of cardinality in children's early years resort to tasks consisting in counting collections of objects, such as the Give-N task (e.g. Condry & Spelke, 2008; Izard, Streri, & Spelke, 2014; Sarnecka & Carey, 2008). Thus, the use of collections appears to emphasize the cardinality of numbers.

Weights. In daily-life, weight is often seen as a property of some definite element, and as such it has no ontological order. For instance, when considering the weight of a stack of dictionaries, it does not matter which one is on top and which one is at the bottom of the stack: computing the total weight simply requires adding the weight of each individual book. Specific weights are assigned to specific entities, and although weight can vary over time, it is rarely encoded along an axis in daily life. Thus, we postulate that weight promotes a cardinal encoding of the situations.

Prices. As with weight, price is usually considered the property of some unordered entities (Gamo et al., 2010). To calculate the total price of a series of items in a store, one must add the individual price of each item, in no specific order. Thus, we believe that prices underline the cardinal nature of numbers.

Experiment 1

The first experiment, a problem sorting task, was designed to assess whether the assumed difference between cardinal and ordinal quantities would lead to different encodings of the problems, reflected by categorization patterns consistent with this distinction. Furthermore, this experiment aimed at validating our choice of materials in order to build upon this cardinal-ordinal dichotomy on subsequent experiments exploring this issue. In other words, it intended to provide evidence for the existence of a fundamental distinction between cardinal and ordinal encodings derived from problems sharing an identical deep structure but differing in regard to the quantities used in their problem statements. Indeed, we expected participants' categories to reveal a difference between the respective encodings of cardinal and ordinal problem statements. The experimental design was modeled on the work from Chi, Feltovich and Glaser (1981), who performed a series of sorting experiments with experts and novices. They showed that, when asked to sort physics problems, novices put together problems sharing similar surface features (e.g. problems featuring pulleys), whereas participants with a higher proficiency in physics favored the use of abstract physics principles (e.g. problems that can be solved using Newton's second law) to sort the problems.

Our prediction regarding such problem statements was that the adults' understanding of the problems would neither be strictly limited to surface features, as was the case for Chi et al.'s (1981) lay participants, nor would it be guided by the mathematical deep structure of the problems. Rather, we hypothesized that participants would sort problems depending on how they encode them. We predicted that participants would tend to group together problems evoking a cardinal encoding (collection, price and weight problems) and that they would group together problems evoking an ordinal encoding (duration, height and number of floors problems), regardless of the problems' surface features.

Method

Participants. In our first experiment, 85 adults participated after giving informed consent (54 women and 31 men, $M = 24.31$ years, $SD = 8.33$). They were recruited from the Paris region on a voluntary basis and they spoke French fluently. None had previously participated in any similar experiment.

Materials and procedure. We created 12 problems: 6 involved cardinal quantities, as previously described (2 collection problems, 2 price problems and 2 weight problems) and 6 featured ordinal quantities, as defined hereinabove (2 duration problems, 2 height problem and 2 floors problems). The problems were written in French (original materials for all experiments are available online at https://osf.io/kz6gh/?view_only=2f3fb6b910844e238ae58fef3c61168a). English translation is provided in Tables 2 and 3). Although the problems all shared the same mathematical structure as those used by Gamo et al. (2010), they differed in that we made sure they all had the exact same number of sentences, and the different values were systematically mentioned in the same order across problems. Cardinal and ordinal problems did not significantly differ in number of words; $t(10) = 0.98$, $p = .35$, independent t-test. The goal was to create cardinal and ordinal problems as comparable to each other as possible. The problems were isomorphic, and the numbers used were randomized across problems. The computation of the problems' solving algorithms only required basic mathematical knowledge (additions and subtractions of values below 15) so that the performance differences could not be explained by computational difficulty or lack of knowledge about the arithmetic operations involved.

In this experiment, each of the 12 word problems was printed on a separate card. As in Chi et al.'s (1981) experiment, the task was to sort the problems into groups based on similarities of solving strategies. The following instructions were given to the participants:

Here are 12 arithmetic word problems. Please read and study carefully each of them. Your task is to sort the problems into groups based on similarities of solution. You can make as many groups of problems as you deem necessary. This is not a speed test: take your time to read and understand each of these problems. *Translated from French.*

Participants all completed the task in less than an hour.

Table 2 Cardinal problems used in Experiment 1. The numerical values respected the following rule: $z < 4 < x < y < 15$.

Quantity used	Pb. ID	Problem statement
Weight	Pb. A	<p>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 now?</p>
Weight	Pb. B	<p>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. In total, how many kilograms is Lola carrying?</p>
Price	Pb. C	<p>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 omelet with mushrooms, and an apple pie. The apple pie costs z euros less than chocolate cake. How much does the second meal cost?</p>
Price	Pb. D	<p>In the stationery shop, Antoine 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 Antoine, and an eraser. The eraser costs z euros less than the ruler. How much will Julie have to pay?</p>
Collection	Pb. E	<p>Paul has x red marbles. He also has blue marbles. In total, Paul has y marbles. Charlene has as many blue marbles as Paul, and some green marbles. She has z green marbles less than Tom has red marbles. How many marbles does Charlene have?</p>
Collection	Pb. F	<p>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?</p>

Table 3 Ordinal problems used in Experiment 1. The numerical values respected the following rule: $z < 4 < x < y < 15$.

Quantity used	Pb. ID	Problem statement
Duration	Pb. G	<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	Pb. H	<p>Sophie travels for x hours. Her trip started during the day. Sophie arrives at y. Fred leaves at the same time as Sophie. Fred's trip lasts for z hours less than Sophie's. What time is it when Fred arrives?</p>
Height	Pb. I	<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	Pb. J	<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>
Floors	Pb. K	<p>Naomi takes the elevator and goes up x floors. She left from the floor where her grandparents live. She arrives to the yth floor. Her brother Derek also takes the elevator from their grandparents' floor. He goes up z floors less than Naomi. What floor does Derek arrive to?</p>
Floors	Pb. L	<p>Karen takes the elevator and goes up x floors. She left from the floor where the gym is. She arrives to the yth floor. Yohan also takes the elevator from the floor where the gym is. He goes up z floors less than Karen. What floor does Yohan arrive to?</p>

Results

The categories created by each participant were coded with a co-occurrence matrix describing how many times two problems were sorted together within the same category. A proximity matrix was then elaborated based on the co-occurrence matrix, describing the average perceived proximity between each problem (see Fig. 3). This matrix specifies which problems were sorted together most frequently; the higher the value between two problems, the higher the proportion of participants who considered these two problems similar. The perceived proximity between any pair of problems whose hypothesized encoding is similar (ordinal-ordinal or cardinal-cardinal) was systematically higher than the perceived similarity between any pair of problems whose hypothesized encoding is dissimilar (cardinal-ordinal).

		Weight problems		Cardinal problems		Collection problems		Duration problems		Ordinal problems		Floors problems		
		Pb. A	Pb. B	Pb. C	Pb. D	Pb. E	Pb. F	Pb. G	Pb. H	Pb. I	Pb. J	Pb. K	Pb. L	
Cardinal problems	Weight problems	Pb. A	-	0.95	0.90	0.89	0.88	0.86	0.79	0.76	0.78	0.78	0.75	0.76
		Pb. B	0.95	-	0.87	0.94	0.90	0.89	0.75	0.74	0.75	0.75	0.72	0.72
	Price problems	Pb. C	0.90	0.87	-	0.91	0.90	0.86	0.76	0.75	0.77	0.77	0.77	0.78
		Pb. D	0.89	0.94	0.91	-	0.87	0.89	0.75	0.76	0.76	0.74	0.74	0.74
	Collection problems	Pb. E	0.88	0.90	0.90	0.87	-	0.95	0.74	0.71	0.74	0.74	0.74	0.75
		Pb. F	0.86	0.89	0.86	0.89	0.95	-	0.75	0.73	0.76	0.72	0.74	0.74
Ordinal problems	Duration problems	Pb. G	0.79	0.75	0.76	0.75	0.74	0.75	-	0.95	0.86	0.87	0.89	0.89
		Pb. H	0.76	0.74	0.75	0.76	0.71	0.73	0.95	-	0.87	0.88	0.91	0.91
	Height problems	Pb. I	0.78	0.75	0.77	0.76	0.74	0.76	0.86	0.87	-	0.95	0.92	0.93
		Pb. J	0.78	0.75	0.77	0.74	0.74	0.72	0.87	0.88	0.95	-	0.91	0.91
	Floors problems	Pb. K	0.75	0.72	0.77	0.74	0.74	0.74	0.89	0.91	0.92	0.91	-	1.00
		Pb. L	0.76	0.72	0.78	0.74	0.75	0.74	0.89	0.91	0.93	0.91	1.00	-

Fig. 3 Co-occurrence proximity matrix between problems. A higher number (and darker hue) indicates a higher co-occurrence frequency.

In order to get confirmatory evidence, we undertook a hierarchical cluster analysis displaying the global taxonomy of the collected categories, using the R package *pvclust* (Suzuki & Shimodaira, 2006). Fig. 4 details the clusters appearing in the dataset. For each cluster, the Approximately Unbiased (AU) p-value can be interpreted as follows: if $AU > .95$, the hypothesis that “the cluster does not exist” can be rejected at the significance level of .05. This suggests that these clusters do not reflect clustering noise and may be observed in a stable manner.

The problems which shared the same quantities were grouped in the same lower clusters. This was expected since they both shared quantities attached to the same world semantics (either cardinal or ordinal) as well as a certain degree of surface. All the problems sharing the same quantities (e.g. the two weight problems, the two duration problems, etc.) were grouped in the same lower clusters. This was expected since they both evoked the same encoding (either cardinal or ordinal) and shared a certain degree of surface similarity as well (two problems involving objects being weighed, for example, will both use words such as “weight”, “scale” or “weighs”, increasing the number of identical surface features). Yet, the two higher clusters

displaying a significant AU p-value respectively regrouped all the ordinal problems and all the cardinal problems, regardless of their more specific surface features. This result supports the hypothesis that the cardinal versus ordinal distinction guided participants' sorting patterns. Thus, as hypothesized, some aspects of the problem statements that pertain neither to the most specific surface features nor to the deep structure of the problems influenced the classification. The Bootstrap probability (BP) values, although presumably more biased (Suzuki & Shimodaira, 2006), confirmed this pattern in the data.

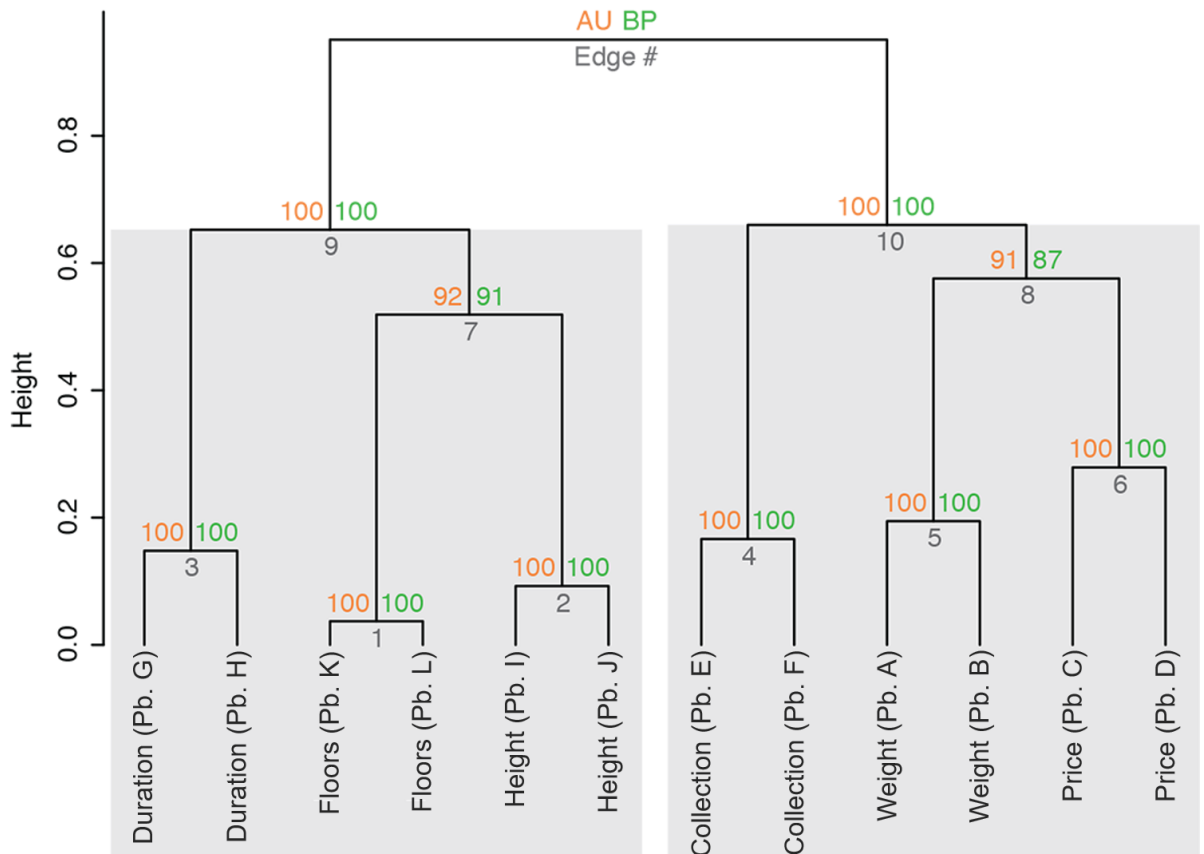


Fig. 4 Cluster dendrogram of the problems. Values are Approximately Unbiased (AU) p-values (Red, left) computed by multiscale bootstrap resampling, Bootstrap Probability (BP) values (green, right) computed by normal bootstrap resampling, and cluster labels (grey, bottom). Clusters with AU ≥ 95 are highlighted by the red rectangles and are considered to be strongly supported by the data. Number of bootstrap samples = 10,000.

Discussion

This experiment sought to establish that participants' categories were guided by the semantic distinction we introduced between cardinal and ordinal quantities. As predicted, the results showed that participants' categories were not just based on the most superficial similarities between problems. Had it been the case, duration problems would not have been considered as closer from number of floors problems

than from collection problems, for instance. Instead, participants perceived the similarity between problems fostering a similar encoding: problems evoking sets of unordered elements were grouped together, and so were problems with elements that could be ordered along an oriented axis. Additionally, participants did not perceive that all problems were isomorphic, since that would have resulted in one category containing all the problems. Participants here did not rely on the deep structure of the problems to make their categories. Instead, the results supported our hypothesis that weight, price and collection problems are encoded differently than height, duration and floors problems, thus substantiating our selection of those two sets of quantities. Here, the cardinal versus ordinal distinction significantly guided the sorting task, thus corroborating the fundamental role of this distinction.

Experiment 2

To evaluate how cardinality and ordinality interact with one another when different situations are being compared, and to gather converging evidence regarding the influence that this distinction holds on adults' apprehension of numerical situations, our second experiment focused on participants' interpretation of different problems as analogous. Participants were presented with an unsolved word problem and asked to determine whether a series of target word problems could be solved analogously. We tested the hypothesis that participants can more easily perceive an analogy between two isomorphic problems if they feature quantities evoking a similar encoding (two problems emphasizing the cardinal nature of numbers, or two problems emphasizing the ordinal nature of numbers) than if they do not.

Methods

Participants. A total of 191 adults participated in this experiment after giving informed consent: 116 women and 75 men, $M = 27.3$ years, $SD = 11.9$. They were recruited from the Paris region on a voluntary basis and spoke French fluently. None had previously participated in any similar experiment.

Materials and procedure. In this experiment, we used the same problems as those created for Experiment 1, with the addition of one cardinal problem and one ordinal problem. Each participant was given a 4-page booklet. On the first page, the instructions read:

Below is an arithmetic word problem. Please read it and then study carefully

the problems presented on the next page. Note, for each of them, if they can be solved using the same solving principle as the problem presented below. This is not a speed test: take your time to read and understand each of these problems. *Translated from French.*

On the same page, a problem statement was printed (“problem A”). This problem was either a cardinal problem (a collection problem) or an ordinal one (a duration problem). On the following page, 6 target problems were presented: 3 ordinal problems (duration, height, floors), and 3 cardinal problems (collection, price, weight). Next to each problem, the participants had to circle their answer “yes” or “no” to the question “can this problem be solved similarly to problem A?”. The next two pages had the same setup with different problem statements, where “problem A” was replaced by “problem B”, that was either a collection problem (if “problem A” had been a duration problem) or a duration problem (if “problem A” had been a collection problem), and six new target problems. The order of problems A and B was randomized between booklets, as was the order in which the target problems were presented. Participants were not given additional paper to write on, to discourage them from engaging in the resolution of every problem before making their choice. Participants all completed the task in less than an hour.

We predicted that the participants’ answers would depend on the similarity between the type of representations fostered by the source and target problems. Namely, participants should perceive the analogy between two cardinal problems or between two ordinal problems more easily than between a cardinal and an ordinal problem. This should translate into an interaction between the cardinal versus ordinal nature of the source problem and that of the target problems.

Results

We computed the rate of detection of similarity between the source and the target problems, depending on the nature of the quantities involved (see Fig. 5). A two-way repeated measures ANOVA was conducted on the rate of perceived similarity with nature of the quantity in the source problem (ordinal or cardinal) and nature of the quantity in the target problem (ordinal or cardinal) as within factors. As expected, there was no main effect of the cardinal or ordinal semantics attached to the source problem ($F(1,190) = 1.07, p = .30, \eta_p^2 = .01$) nor of the semantics attached to the target problem ($F(1,190) = 0.10, p = .08, \eta_p^2 < .01$). There was, however, a significant interaction effect between the two factors, indicating that cardinal target problems

were more likely to be selected by the participants when a cardinal source problem was presented, and that ordinal target problems were more likely to be chosen when the source problem was ordinal as well ($F(1, 190) = 72.20, p < .001, \eta_p^2 = .28$). In addition to the interaction, we performed 2-by-2 comparisons of the rate of perceived similarity between cardinal and ordinal target problems, depending on the semantics imbued in the source problem. Results showed that cardinal target problems were judged analogous to cardinal source problems significantly more often than ordinal target problems (82.4% for cardinal-to-cardinal analogy; 60.9% for cardinal-to-ordinal analogy; $t(190) = 6.50, p < .001, \eta_p^2 = .18$, paired t-test). Similarly, ordinal target problems were judged analogous to ordinal source problems significantly more often than cardinal target problems (84.8% for ordinal-to-ordinal analogy; 62.1% for ordinal-to-cardinal analogy; $t(190) = 7.25, p < .001, \eta_p^2 = .22$, paired t-test). In other words, participants identified collection source problems as analogous to other cardinal problems more frequently than they did to ordinal target problems. Reciprocally, the duration source problems were more frequently perceived as analogous to other ordinal problems than to cardinal problems.

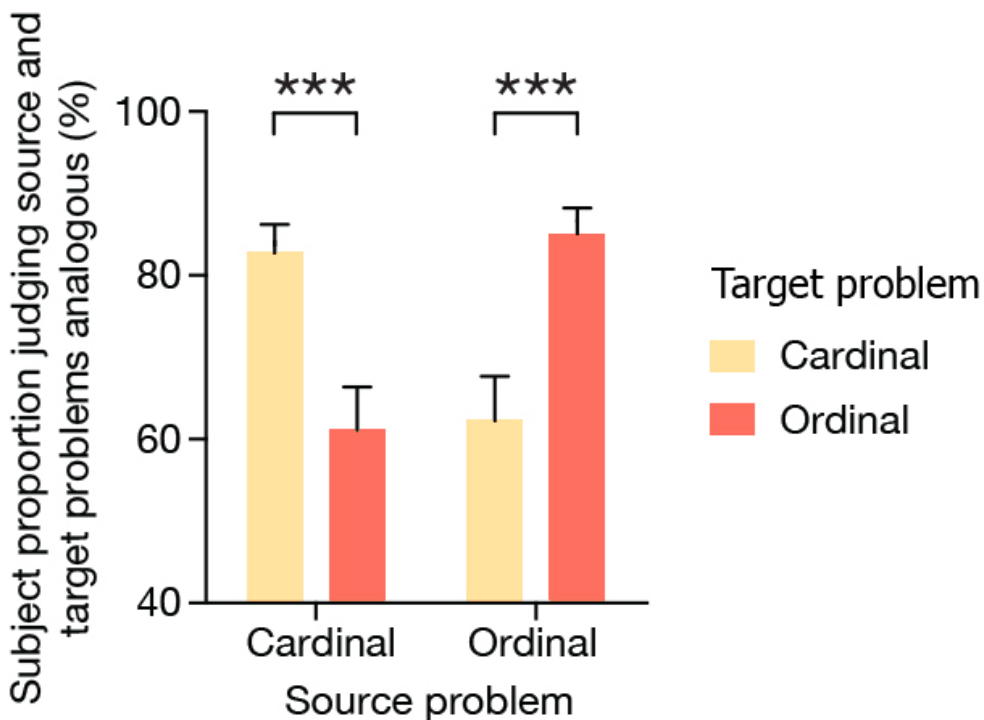


Fig. 5 Rate of perception of the analogy between the source and target problems, depending on the nature of their quantities. Vertical bars denote .95 confidence intervals. *** $p < .001$, paired t-test.

Overall, the analogy rate was higher when the source and target problems both evoked a similar encoding (either two cardinal encodings or two ordinal encodings) than when they evoked dissimilar encodings (a cardinal encoding and an ordinal encoding). Our hypothesis regarding the influence of the encoding on the detection of similarity of solving pattern was thus supported: participants' decisions were significantly influenced by the semantics attached to the quantities used in the problems.

Discussion

With this experiment, we showed that the encoding difference between cardinal and ordinal problems influences the perception of analogies between isomorphic problems. Participants had significantly more difficulties identifying that two problems shared the same solving principle when these problems featured elements evoking different aspects of their knowledge about the world. This corroborates our claim that general abstract semantic properties, such as cardinality or ordinality, play a crucial role in the encoding of mathematical word problems. Depending on the semantics evoked by the problems, participants encode different representations. The representations in turn limit or foster participants' identification of the analogous relations between the problems. Our results support the idea that participants encode a representation whose nature depends on the world knowledge evoked by the problem statement.

While the interaction between the cardinal versus ordinal nature of the target and source problems indicates that participants' encoding of the problems was significantly influenced by the quantities they featured, it might be that participants could have overcome this inability to perceive the similarity between cardinal and ordinal problems if they had tried to directly map the structures of the problems to one another. In other words, suppose participants were encouraged to compare one specific target problem with the source problem instead of simply being asked to select the analogous problems among a series of potential candidates. Would they go beyond their initial encoding of the situation and identify the isomorphism between the problems? We designed a third experiment to answer this question.

Experiment 3

Experiment 3 aimed at replicating Experiment 2's findings and assessing their robustness by directly presenting pairs of problems to the participants, instead of asking them to identify among a list of target problems the ones that shared a solution principle with the source problem. We hypothesized that the effect observed in Experiment 2 could be replicated in a one-to-one comparison setting, in which participants are encouraged to directly contrast two problem statements. We assumed that when the source and the target problems featured the same type of quantity – cardinal or ordinal – participants would acknowledge the solution equivalence more often than when problems featured different types of quantity.

Methods

Participants. A total of 147 adults participated in this experiment after giving informed consent: 60 women and 87 men, $M = 30.0$ years, $SD = 11.5$. They were recruited on a voluntary basis through social networks and by emails. All participants spoke French fluently and none had previously participated in any similar experiment. Five participants were removed since they failed to provide any answer to one or more questions in the experiment. The analyses bear on the 142 remaining participants (59 women and 83 men, $M = 29.1$ years, $SD = 10.2$ years). Participants all completed the task in less than an hour.

Materials and procedure. This experiment was conducted on the Qualtrics platform for online experiments. The source and target problems were the same as those used in Experiment 2. On the first page, the instructions read:

Below is an arithmetic word problem. Please read it carefully. On the next pages, you will be presented with a series of arithmetic problems. Indicate, for each new problem, whether it can be solved with the same solving principle as the problem presented below. This is not a speed test: take your time to read and understand each of these problems. *Translated from French.*

A source problem was then presented, evoking either a cardinal or an ordinal encoding. The following 6 pages repeated the source problem, and then presented a new problem below. Each time, the following question was displayed: “Can these two problems be solved using a similar solution principle?”. After 6 target problems had been introduced, a new source problem was presented (a cardinal problem if the

first target problem was ordinal, an ordinal problem if the first target problem was cardinal), with the same instructions as before, followed by 6 new target problems, each on an individual screen. The source and target problems were the same as those used in Experiment 2. Participants all completed the task in less than an hour.

Results

For each type of source problem, we analyzed the percentage of participants answering that the source and target problems might be solved following a similar solution principle (see Fig. 6). A two-way repeated measures ANOVA was conducted on participants' rate of identified similarity, with nature of the quantity in the source problem (cardinal or ordinal) and nature of the quantity in the target problem (cardinal or ordinal) as within factors. As in Experiment 2, there was no main effect of the cardinal or ordinal semantics attached to the source problem ($F(1,140) = 2.14$, $p = .15$, $\eta_p^2 = .02$) nor of the semantics attached to the target problem ($F(1,140) = 0.41$, $p = .53$, $\eta_p^2 < .01$). However, we replicated the interaction observed in Experiment 2 between the semantic nature of the source problems and that of the target problems ($F(1, 140) = 73.39$, $p < .001$, $\eta_p^2 = .34$).

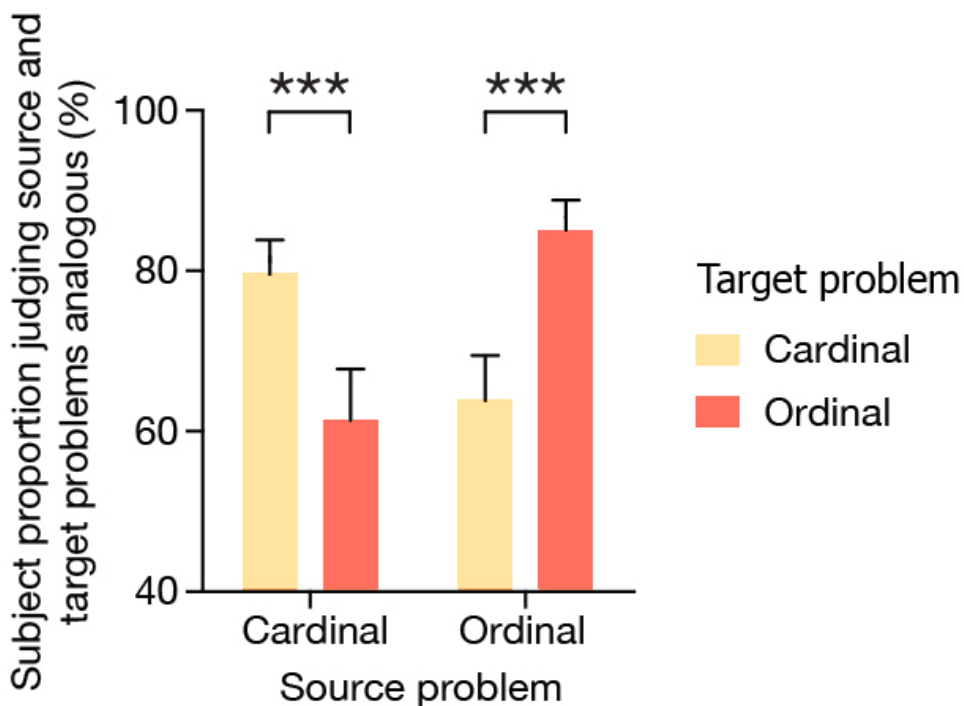


Fig. 6 Rate of perception of the analogy between the source and target problems, depending on the nature of their quantities. Vertical bars denote .95 confidence intervals. *** $p < .001$, paired t-test.

In addition to the interaction, 2-by-2 analyses revealed that cardinal target problems were judged analogous to cardinal source problems significantly more often than ordinal target problems (80.9% for cardinal-to-cardinal analogy; 62.6% for cardinal-to-ordinal analogy; $t(140) = 5.51$, $p < .001$, $\eta_p^2 = .18$, paired t-test). Similarly, ordinal target problems were judged analogous to ordinal source problems significantly more often than cardinal target problems (87.7% for ordinal-to-ordinal analogy; 65.7% for ordinal-to-cardinal analogy; $t(140) = 6.91$, $p < .001$, $\eta_p^2 = .25$, paired t-test). In other words, despite being encouraged to directly compare one source problem with one target problem, participants' ability to perceive the analogy between the two problems was still dependent on the semantics they carried.

Discussion

In this experiment, it was shown that the effect of the distinction between cardinal and ordinal quantities on the encoding of the situations described in the problems was not altered by the direct presentation of two problems side by side. As predicted, participants had significantly more difficulties identifying that ordinal-cardinal pairs of problems could be solved in the same way, even though the simultaneous presentation of both problems should have facilitated their mapping. Interestingly, the results of Experiment 2 were replicated in this new experimental setting. Our hypothesis regarding the influence of the cardinal versus ordinal distinction on the detection of similarity of solving pattern was thus supported by both experiments. The question that follows from these two experiments regards the role of these robust encoding mechanisms in the choice of a solving algorithm. To what extent do the constructed representations dictate participants' solving strategies? By using problems designed to evoke cardinal, ordinal, or hybrid encodings, the next experiment was designed to evaluate how these encoding differences may influence participants' choice of an algorithm in a more traditional solving task.

Experiment 4

The goal of this fourth experiment was twofold. First, we intended to demonstrate that participants' ability to use a specific solving strategy directly depended on the nature of their semantic encoding of the problem. Second, we aimed at showing that by changing the semantics imbued in a problem statement, we could significantly alter its encoding on the cardinal versus ordinal dimension. To this end, we

introduced a new type of problems, in addition to the cardinal and ordinal problems used in the previous experiments. We called those new problems hybrid because they were designed to be less cardinal than the cardinal problems presented in Experiments 1, 2 and 3, while simultaneously being less ordinal than the ordinal. The goal of this fourth experiment was twofold. First, we intended to demonstrate that participants' ability to use a specific solving strategy directly depended on the nature of their semantic encoding of the problem. Second, we aimed at showing that by changing the semantics imbued in a problem statement, we could significantly alter its encoding on the cardinal versus ordinal dimension. To this end, we introduced a new type of problems, in addition to the cardinal and ordinal problems used in the previous experiments. We called those new problems "hybrid", as they were meant to elicit an ordinal encoding using cardinal quantities. Indeed, those hybrid problems involved only cardinal quantities (price, weight, collection) but their problem statements featured a scenario fostering an ordinal encoding by describing how the cardinal quantities changed over time. For instance, hybrid weight problems were created by describing the weight of a baby growing over time in order to favor an ordinal representation of the weight. We predicted that introducing those characteristics without changing the quantities themselves would influence the encoding of the problems and the algorithms subsequently implemented.

Since we instructed participants to solve the problems using as few operations as possible, our main hypothesis regarded the rate of use of the 1-step algorithm. We predicted that problems involving ordinal quantities would lead to a greater use of the 1-step algorithm than problems involving cardinal quantities, due to the ordinal encoding making it easier to perceive the validity of this solution. Second, we hypothesized that hybrid problems would lead to a significantly higher rate of 1-step algorithm than cardinal problems, due to the ordinal semantics attached to the problem statements. Additionally, we aimed at assessing whether hybrid problems would be solved by the 1-step algorithm as often as ordinal problems, or not.

Methods

Participants. Participants were students from a second-year university psychology class at the University of Bourgogne. They participated in exchange for course credit. A total of 181 students participated in this experiment after giving informed consent (123 women and 58 men, $M = 23.35$ years, $SD = 7.82$). All the participants spoke French fluently. None had previously participated in any similar experiment.

Materials and procedure. A pool of 18 word problems was used for this experiment: the same 12 problems (6 cardinal and 6 ordinal) as in the two first experiments, and 6 new hybrid problems (see Table 4). Each participant saw 9 problems in total: 3 of each category.

Table 4 Example of hybrid problems. The numerical values respected the following rule: $z < 4 < x < y < 15$.

Quantity used	Hybrid problem statement
Weight	<p>During his first year, David gained x kilograms. At birth, he already weighed a certain weight. After a year, David weighs y kilograms. At birth, David and Lara had the same weight. During her first year, Lara gained z kilograms less than David did. How much does Lara weigh after one year?</p>
Price	<p>For Christmas, Rachel got x euros. She already had some money. Now she has y euros in total. Before Christmas, Zoe had as much money as Rachel. For Christmas, Zoe got z euros less than Rachel did. How much money does Zoe have now?</p>
Collection	<p>During the afternoon, Patricia catches x fish. She puts those fish in her basket, with the other fish she caught during the morning. By the end of the day, Patricia has y fish in her basket. During the morning, Arthur caught as many fish as Patricia did. During the afternoon, Arthur catches z fish less than Patricia does. In total, how many fish did Arthur catch today?</p>

The participants all received 10-page booklets with instructions printed on the first page. The instructions read:

You will find an arithmetic problem on each page of this booklet. Your task is to solve the problems using as few operations as possible. You can use the ‘draft’ area, but please copy in the ‘response’ area all the operations that you used to come up with the solution. This is not a speed test: take your time to read and understand each of these problems. Remember that the goal is to solve the problems using as few operations as possible. For every problem, we ask you to write down every operation(s) that you used to come up with the solution, even the simplest one that you can calculate mentally. For instance, the computation “ $15 - 6 - 2 = 7$ ”, should not be written as a unique operation, but broken down as “ $15 - 6 = 9$ ” and “ $9 - 2 = 7$ ”, which then counts

for two operations. *Translated from French.*

Problem order was randomized across booklets. Each page in the booklet was divided in three parts: the problem statement, the “draft” area and the “response” area. Participants all completed the task in less than an hour.

Scoring. A problem was considered as correctly solved when the obtained result came with the appropriate calculations. The strategies leading to success were categorized either as correct 1-step algorithm or as correct 3-step algorithm. 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. Other answers were considered as false.

Results

The percentage of correct solving using both algorithms was calculated. Fig. 6 details the rate of use of the 1-step algorithm for each problem category. A one-way repeated measures ANOVA was conducted on the rate of use of the 1-step algorithm between cardinal, hybrid and ordinal problems. Results indicated that there was a main effect of problem category (cardinal/hybrid/ordinal) ($F(2,360) = 52.13, p < .001, \eta_p^2 = .22$). Three paired sample t-tests were used to compare the different conditions, with Bonferroni adjustment for multiple comparisons. As predicted, cardinal problems led to a significantly lower rate of 1-step algorithm ($M = 0.28, SD = 0.39$) than ordinal problems ($M = 0.47, SD = 0.39$); ($t(180) = 9.34, p < .001, d = 0.48$), which supported the hypothesis that the use of cardinal versus ordinal quantities significantly influenced the encoding of the problem statements, and subsequently shaped the solving algorithms used by the participants. Moreover, as hypothesized, the 1-step algorithm was more frequently used on hybrid problems ($M = 0.38, SD = 0.40$) than on cardinal problems; ($t(180) = 5.91, p < .001, d = 0.24$). This result showed that, by giving specific semantic properties to a problem statement, it was possible to influence the encoding and manipulate which solving algorithm participants would use. A semantically cardinal quantity presented in an ordinal context could thus lead more often to the encoding of an ordinal representation than a cardinal quantity presented in a context that does not present ordinal features.

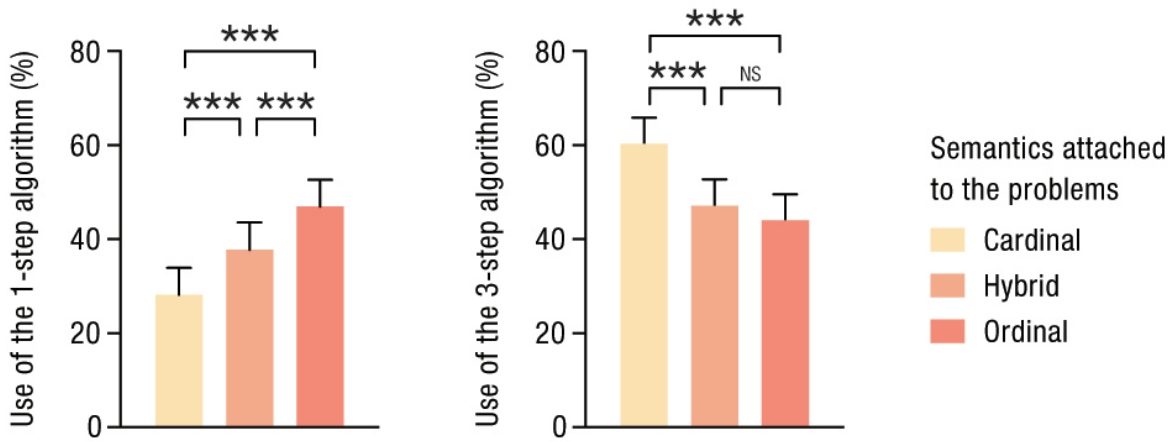


Fig. 7 Rate of use of the 1-step algorithm, depending on the world semantics. Vertical bars denote .95 confidence intervals. *** $p < .001$ (paired t-test).

Interestingly, the comparison between hybrid problems and problems with ordinal quantities revealed that the rate of use of the 1-step algorithm was still higher on ordinal problems ($M = 0.47$, $SD = 0.39$); $t(180) = 4.93$, $p < .001$, $d = 0.23$. This seems to indicate that, while the manipulation that was performed to “ordinalize” cardinal quantities had a significant impact when compared to standard cardinal problems, an “ordinalized” quantity remained less ordinal than a typical ordinal one.

Regarding the rate of use of the 3-step algorithm, we performed a one-way repeated measures ANOVA to evaluate if it differed between cardinal, hybrid and ordinal problems. Results indicated the presence of a main effect of problem category (cardinal/hybrid/ordinal) on the use of this algorithm as well ($F(2,360) = 27.80$, $p < .001$, $\eta_p^2 = .13$). We used paired sample t-tests to perform pairwise comparisons between the three conditions, with Bonferroni adjustment for multiple comparisons. Results showed that participants resorted to the 3-step algorithm more often on cardinal problems ($M = 0.60$ $SD = 0.39$) than on ordinal problems ($M = 0.44$ $SD = 0.38$); $t(180) = 6.79$, $p < .001$. Interestingly, they also used the 3-step algorithm more often on cardinal than on hybrid problems ($M = 0.47$ $SD = 0.38$); $t(180) = 5.80$, $p < .001$. This is in line with the idea that introducing cardinal quantities in an ordinal context could help “ordinalize” participants’ representation. There was, however, no significant difference between the rate of use of the 3-step algorithm on hybrid and on ordinal problems; $t(180) = 1.38$, $p = .17$.

Discussion

In this experiment, the analysis of the solving algorithms provided cues on how the encoding of the problems influences participants’ solving strategies. Participants’ use

of the shortest algorithm was dependent on the type of quantities involved in the problem statement. Despite participants being explicitly instructed to use as few operations as possible to solve the problems, they struggled to find the 1-step algorithm on cardinal problems, and they tended to use the 3-step algorithm to solve these problems instead. This experiment further supported the claim that the ordinal versus cardinal dimension was the main factor constraining algorithm choice. Indeed, a change in strategy choice followed the “ordinalization” of cardinal quantities, thus showing that the ability to use the 1-step algorithm was directly dependent on how much the problem statement emphasized the ordinal nature of its numerical values.

So far, we have studied how arithmetic word problems are initially encoded and how the interpreted representations in turn influence the solving algorithms used. However, as mentioned in the introduction, the initial encoding of a problem does not always provide a solution to the solver. We then investigated whether participants can overcome their initial representation of the problems when the one they first encoded does not lead to a solution. That is, we created the conditions to explore the difference between situations in which one has to construct a new representation of the situation presented, and those in which there is no need for a new representation to be constructed. Such a recoding process would imply to disregard the cardinal semantics evoked by specific quantities, and to construct a new encoding of the situation regardless of their influence.

Experiment 5

In this fifth experiment, we designed situations meant to have participants construct a representation leading to a dead end. In other words, we created problems for which the initial encoding would not provide a successful solving algorithm, thus tempering with the solving process. The problems could nonetheless be solved if participants constructed a different encoding of the situation. We tested the participants' proficiency to use the 1-step solving algorithm on problems that would spontaneously elicit the 3-step algorithm. For that purpose, cardinal problems that only featured two numerical values were introduced, making the 3-step solving algorithm impossible to use. By contrast, the 1-step algorithm was still efficient for reaching the solution with the two remaining numerical values.

A solvability judgment task requested participants to tell whether problems were solvable and to write down the solution of the solvable problems. Our aim was

to show that semantic constraints associated with cardinal situations would lead adult participants to evoke an encoding incompatible with the 1-step algorithm, leading them to fail to find the solution to the problems. We expected them to incorrectly dismiss a perfectly valid solving algorithm, erroneously labeling a 1-step subtraction problem as unsolvable. We also expected that succeeding in solving cardinal problems would require an extra representational step, since the initial representation favors the use of the 3-step algorithm. We assumed that the construction of a new representation would be costly and time consuming. Therefore, our predictions were twofold: first, we hypothesized that participants would incorrectly reject the cardinal solvable problems more often than the ordinal solvable problems because of the conflict between a cardinal representation and the 1-step algorithm. Second, when correctly solved, cardinal problems would require a significantly longer response time than correctly solved ordinal problems, because of the extra step needed to build a new representation compatible with the 1-step algorithm.

Methods

Participants. This experiment was conducted online, on the survey platform Qualtrics. Based on the difference in the rate of use of the 1-step algorithm on cardinal and ordinal problems in Experiment 4, we determined sample size with uncertainty and publication bias correction using the *ss.powed.dt* function from the *BUCSS R* package (Anderson, Kelley, & Maxwell, 2017). With a desired level of statistical power of .9 and a desired level of assurance of .9, the minimum sample size was estimated at 33 participants. Survey link was sent through social networks only. We decided to keep the survey open for one week before assessing if the target sample size had been reached. After one week, a total of 89 adults had participated voluntarily (50 women and 39 men, $M = 32.1$ years, $SD = 13.4$ years). All participants spoke French fluently and none had previously participated in any similar experiment. Because part of the analyses were performed on response times, we removed 15 participants who either mentioned taking a break during the test or who answered at least one of the questions in less than 5 seconds (which meant they either mis-clicked or did not take the time to read the problem). The analyses bear on the 74 remaining participants (44 women and 30 men, $M = 33.8$ years, $SD = 13.4$ years).

Materials and procedure. The problems used in this experiment were similar to the ones in previous experiments, except for the value of *Part 1* (see Fig. 1) that was removed from the statements so that the 3-step algorithm could not be used anymore.

Consequently, the only way to solve the problems was to resort to the 1-step algorithm, which required using the remaining values of *Whole 1* and of the *Difference* (see Fig. 1). Table 5 presents 6 examples of such problems (3 cardinal and 3 ordinal problem statements) created from the ones used in the previous experiments by removing the x value corresponding to *Part 1*. Ordinal problems were 333.5 characters long on average (SD = 38.37) and cardinal problems were 304 characters long on average (SD = 44.94). This length difference was not statistically significant ($t(10) = 1.18, p = .26$, paired t-test).

Table 5 Example of target problems used in the study. Changes introduced from the problems in experiments 1 to 3 are italicized in the table for the sake of clarity, but they were not made apparent in the experiment. Translated from French.

Cardinal target problems	Ordinal target problems
<p>Paul has <i>a certain amount of</i> red marbles. He also has blue marbles. In total, Paul has 14 marbles. Jolene has as many blue marbles as Paul, and some green marbles. She has 2 green marbles less than Paul has red marbles. How many marbles does Jolene have?</p>	<p>Sofia travelled <i>for a certain time</i>. Her trip started during the day. Sofia arrived at 14 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?</p>
<p>In the store, Anthony wants to buy a ruler <i>costing a certain price</i>. He also wants a notebook. In total, that will cost him 14 dollars. Julie wants to buy the same notebook as Anthony, and an eraser. The eraser costs 2 dollars less than the ruler. How much will Julie have to pay?</p>	<p>Slouchy Smurf is <i>a certain height</i>. He climbs on a smurf table. He now attains the height of 14 centimeters. Grouchy Smurf climbs on the same table as Slouchy Smurf. Grouchy Smurf is 2 centimeters shorter than Slouchy Smurf. What height does Grouchy Smurf attain when he climbs on the table?</p>
<p>Joe takes a Russian dictionary weighing <i>a certain weight</i>. He also takes a Spanish dictionary. In total, he is carrying 14 kilograms of books. Lucy takes Joe's Spanish dictionary and a German dictionary. The German dictionary weighs 2 kilograms less than the Russian dictionary. How many kilograms is Lucy carrying?</p>	<p>Katherine took the elevator and went up <i>a certain number of floors</i>. She left from the floor where the gym is. She arrived to the 14th floor. Yohan also took the elevator from the floor where the gym is. He went up 2 floors less than Katherine. What floor did Yohan arrive to?</p>

Although our predictions only regarded solvable problems, we also included unsolvable fillers in the materials, so that not every problem had a solution. Among those fillers the value of *Part 1* was preserved, and the value of *Whole 1* was removed

instead, which made the problems unsolvable with either algorithm. Thus, an equal number of fillers was introduced to achieve a uniform distribution of solvable/unsolvable answers. Problem order and numerical values were randomized between participants. On the first page of the online experiment, the following instructions were written:

You will find an arithmetic problem on each page of this survey. Your task is to identify which problems can be solved and to indicate for each of them the operation you used to solve it, as well as the solution you found. Be careful: some of the problems cannot be solved with the available information, thus your answer in such cases should be ‘it is not possible to find the solution’. This is not a speed test: take your time to read and understand each of these problems. *Translated from French.*

On each page of the survey, a problem was displayed with the following question below it “Given the data provided, is it possible to find the solution?” and two buttons “Yes” and “No”. When the participants pressed “Yes”, two new questions appeared, asking them to indicate respectively the operation needed to solve the problem and the result of the operation. Participants used the keyboard to write down their answers. After participants answered all 12 problems, a new page was displayed asking them for their gender, date of birth, and whether they made any breaks during the completion of the experiment. Participants all completed the task in less than an hour.

As previously stated, our first prediction was that participants would perform better on solvable problems with ordinal quantities compared to solvable problems with cardinal quantities. Indeed, we believed that problems whose spontaneous representation was associated with a 3-step algorithm would often lead participants to ignore the 1-step algorithm, due to cardinal representations being incompatible with the shortest algorithm. Our second prediction regarded the cardinal problems that were correctly solved by the participants despite the conflict between a cardinal encoding and the 1-step algorithm. We hypothesized that higher response times would be recorded on successfully solved cardinal problems compared to successfully solved ordinal problems, due to participants needing additional time to overcome their initial encoding of the situations and build a new representation, compatible with the 1-step algorithm.

Results

The dependent variable was the percentage of correct answers on solvable problems. Ordinal solvable problems were successfully solved in 91.9% of the trials, and cardinal solvable problems in 68.5% of the trials (see left graph of Fig. 7). A paired t-test was performed on participants' mean rate of success for cardinal and ordinal problems and showed that the difference was statistically significant ($t(73) = 6.38, p < .001, d = 0.97$), therefore supporting our first hypothesis.

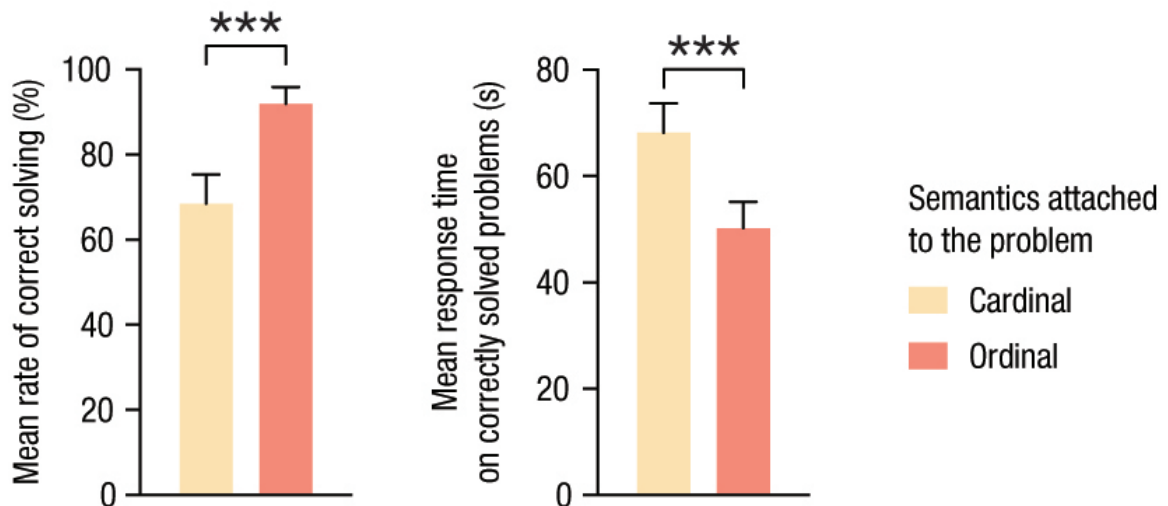


Fig. 8 Mean rate of correct resolution (left) and mean response time on correctly solved problems (right) depending on the semantic nature of the quantities used in the problems. Vertical bars denote .95 confidence intervals. *** $p < .001$ (paired t-test).

Response times on correctly solved cardinal and ordinal problems were then compared in order to test our prediction that accessing the correct 1-step algorithm on problems inducing a cardinal encoding incompatible with this algorithm would require higher response times than it would on problems evoking an ordinal representation. On average, participants took 68.7 seconds to successfully solve cardinal problems, and 49.8 seconds for ordinal problems (see right graph of Fig. 7). Because we only considered the response times for correctly solved problems, the number of measures per participant could vary from 0 to 6, so we resorted to a mixed model analysis instead of a repeated measures ANOVA. We removed 4 participants who did not manage to correctly solve at least one cardinal and one ordinal problems, since no comparison could be made between their response times in both conditions. A linear mixed model with participants as a random factor and problem type (cardinal versus ordinal) as a fixed effect showed that the difference between cardinal and ordinal problems had a significant effect on response times of successfully solved problems ($F(1,69) = 20.38, p < .001$), thus supporting our second hypothesis.

Discussion

As predicted, it was more difficult for participants to use the 1-step algorithm for cardinal than for ordinal problems, despite the 1-step algorithm being the only remaining possibility to find the solution. Besides, as hypothesized, when participants overcame the difficulty and found the solution to a cardinal problem, it required extra processing time, presumably attributable to the construction of a new representation of the situation.

While the fourth experiment showed that the 1-step algorithm is more frequently used on ordinal than on cardinal problems, this fifth experiment showed that this effect is not the result of a mere preference but, instead, seems to be the consequence of strong limitations imposed by the type of quantity used. In fact, this effect was so pervasive that in many cases adult participants failed to see that these one-step subtraction problems could be solved at all. Yet, not all the participants failed, and some of them even managed to use the 1-step algorithm in certain cases. However, in order to overcome the constraints imposed by their world knowledge about the problem's quantities and use a conflicting solving algorithm, the participants had to discard their initial representation and construct a new encoding closer to the problem's mathematical structure. The existence of such a recoding step, akin to a re-representation process (Vicente, Orrantia, & Verschaffel, 2007), was supported by the longer response times required on correctly solved cardinal problems.

Once participants realize that the 3-step algorithm cannot be used given the available information, they might be tempted to discard these problems as unsolvable and move on. In the 6th experiment, we went a step further and provided participants with a potential solution to the problems, one that they would not usually consider. By giving them such a clue, we were able to assess their difficulty to construct an alternate encoding of the situation even when directly incited to do so.

Experiment 6

In this experiment, we provided participants with the 1-step solution algorithm of each problem and asked them to directly evaluate its validity. Because of the high failure rates on cardinal problems in Experiment 5, we tested the bolder hypothesis that providing the solution algorithm would not be sufficient to systematically foster an appropriate encoding of the situation by the participants. We thus hypothesized

that even when explicitly presented with the solution, participants would reject it more often in the cardinal condition than in the ordinal one since their encoding of cardinal problems would conflict with their solution (i.e. promoting an unusable 3-step algorithm instead of the 1-step algorithm). Because of the need to overcome the initial representation, we hypothesized that the correct identification of the solution would require more time for cardinal than for ordinal problems. This experimental paradigm resembled the one used in a recent study we conducted on expert mathematicians (Gros, Sander & Thibaut, 2019). It differed in that here the participants could take as long as they wished to complete the task, whereas in Gros et al. (2019), participants were explicitly told to solve the problems as fast as possible, with the intended purpose of increasing their error rates. The absence of time constraint was meant to give participants the opportunity to read the problems until they were certain of their decision and to engage in a recoding of their initial representation if need be.

Methods

Participants. A total of 223 adults participated in this experiment after giving informed consent. They were recruited through social networks and emails. All spoke French fluently and none had previously participated in any similar experiment. Among them, 27 were removed from the analysis because they either took a break during the test or answered at least one of the questions in less than 5 seconds (which meant they did not take the time to read the problem). The analyses were performed on the remaining 196 participants (135 women and 88 men, $M = 34.5$ years, $SD = 14.8$ years).

Materials and procedure. The only difference between the present experiment and the previous one was the fact that a solution was proposed. Instead of having the participants solve the problems themselves, a solution was proposed for each problem, and they were asked to judge whether the provided solution was valid or whether the problem was unsolvable. For every problem, the question “Given the data provided, is it possible to find the solution?” was displayed. Two choices appeared below: (a) “No, we do not have enough information to solve this problem.” and (b) “Yes: *numerical value 1 – numerical value 2 = result. Sentence presenting the result*”. For instance, on one of the elevator problems, the option (b) was: “Yes: $11 - 2 = 9$. Karin arrives at the 9th floor.” Participants all completed the task in less than an hour.

Results

As in Experiment 5, we first analyzed the ratio of correct answers on solvable problems depending on the type of quantities used. The left graph of Fig. 8 shows that, among the target problems, the cardinal ones had a lower success rate (63.6%) than the ordinal ones (88.4%). A paired t-test performed on the participants' mean rate of success confirmed that this difference was significant ($t(195) = 9.25, p < .001, d = 0.87$).

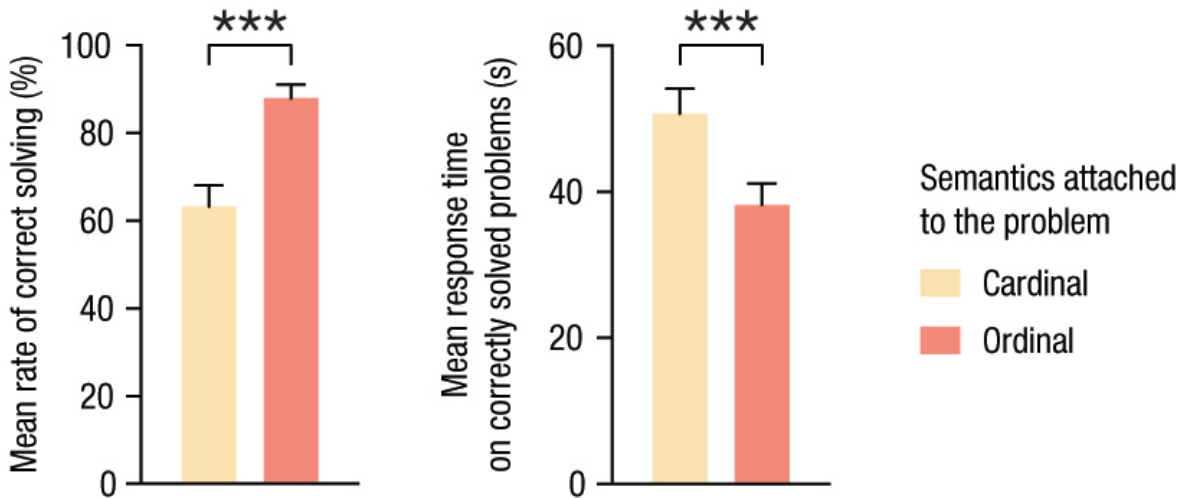


Fig. 9 Mean rate of correct resolution (left) and mean response time on correctly solved problems (right) depending on the semantic nature of the quantities used in the problems. Vertical bars denote .95 confidence intervals. *** $p < .001$ (paired t-test).

In order to assess the validity of our second hypothesis, we analyzed the response times of correct answers on the target solvable problems. The 26 participants who did not manage to correctly respond to at least one cardinal and one ordinal problems were removed from this analysis, since no comparison could be made between their response times in both conditions. The right graph of Fig. 8 shows that providing a correct answer required a shorter response time for ordinal (38.6 seconds) than for cardinal problems (51.4 seconds). A linear mixed-model with participants as a random effect and the cardinal versus ordinal distinction as a fixed factor confirmed that the effect was statistically significant ($F(1,169) = 30.28, p < .001$), supporting the second hypothesis.

Discussion

This experiment, involving a solution validity assessment task, supported the effects observed in the previous one, involving a solution discovery task. The analyses indicated that even when the correct solution was provided, it was more difficult for

them to accept it when it was not compatible with the initial encoding of the problem. Despite the problems being solvable with a mere subtraction, participants went so far as to reject the correct solution and dismiss the problems as “unsolvable”. Furthermore, overcoming this difficulty required more time, thus supporting our prediction of the need for an extra processing step when faced with an inapt representation. These results suggest that the encoding effects identified in the 5th experiment are not restricted to the elaboration of a solving strategy, but also to the evaluation of its validity. This experiment provides additional evidence that the cardinal versus ordinal distinction constrains the encoding of problems, since even when no solving algorithm had to be produced by the participants, their interpretation precluded them from considering the given solution – albeit a single subtraction – as an acceptable one.

General discussion

Taken together, the present six experiments shed light upon the foundational part played by the cardinal versus ordinal distinction in adults’ reasoning about numerical situations. The first experiment validated our choice of materials by showing that the distinction between cardinal and ordinal quantities drives adults’ sorting patterns. Experiments 2 and 3 demonstrated that the encoding difference between cardinal and ordinal problems impacts participants’ success in perceiving problems as analogous, even when explicitly instructed to directly compare two problems. The fourth experiment’s findings were twofold. First, it proved that the distinction between cardinal and ordinal problems influences adults’ choice of a solving algorithm. Second, it showed that it is possible to manipulate a problem’s semantics by presenting cardinal quantities in a context emphasizing the ordinality of the problem’s values. The changes that were introduced to ordinalize the cardinal problems had an effect on participants’ algorithm choice, thus bolstering the importance of the cardinal versus ordinal distinction and strengthening the view that the encoding difference observed depended on the semantic dimensions manipulated in the problems. The fifth experiment showed that even when only one solution was available, adult participants had difficulties to find it when it is not compatible with what was assumed to be their spontaneous encoding of the problems. They were more likely to judge that cardinal problems cannot be solved and, when they did find the solution to a cardinal problem, it nonetheless required a longer reasoning time. Finally, the sixth

experiment demonstrated that presenting the participants with a potential solution did not suppress the effect observed in Experiment 5, as participants continued to struggle to identify the solution of cardinal problems.

The fact that the distinction between cardinal and ordinal quantifications could exert a robust and pervasive effect on adults' apprehension of numerical situations illustrates the foundational nature of this distinction for the human mind. Despite a growing body of research on the development of the notions of cardinality and ordinality in children, especially in learning how to count, little is known regarding the influence of this dimension on adult mathematical reasoning. Our understanding of mathematics is deeply rooted into our understanding of the world (Fischbein, 1987; Hofstadter & Sander, 2013; Lakoff & Nuñez, 2000) and, as a result, we tend to apply real-life constraints to abstract mathematical notions. Here, we showed that our perception of ordinality and cardinality in scenes of our daily-lives has a profound effect on our ability to engage in mathematical reasoning in these situations.

Regarding mathematical problem solving, the range of content effects we have put forward in this paper showcase the fact that even for problems involving relatively elementary arithmetic operations, humans hardly manage to completely disregard context, and their problem-solving skills suffer from this shortcoming. However, the influence of world knowledge on mathematical word problem solving is not surprising considering that it leads to the making of inferences which are correct most of the times, since the surface features of situations are highly correlated with deeper principles, as suggested by the *kind world hypothesis* (Bassok et al., 2008; Blessing & Ross, 1996; Gentner & Medina, 1998; Goldstone, 1994; Trench & Minervino, 2015). Consequently, people tend to rely on those superficial cues which help them find the solution. Difficulties arise when the world knowledge evoked by a problem and its deep structure are semantically incongruent. Individuals are then bound to struggle, and sometimes even fail. This is especially problematic since mathematics education does not usually control for content effects (Bassok et al., 1998; Lee, DeWolf, Bassok, & Holyoak, 2016), which is partly due to mathematics being primarily considered the realm of abstraction (Davis, Hersh, & Marchisotto, 2011; Russell, 1903). Although arithmetic word problems are a central part of mathematics education and teachers are usually encouraged to provide real-world examples to illustrate the notions being taught (e.g. Richland, Stigler, & Holyoak, 2012; Rivet & Krajcik, 2008), the use of concrete examples to teach new notions has also been shown to have a detrimental

effect on transfer (Son & Goldstone, 2009; Goldstone & Sakamoto, 2003; Day, Motz & Goldstone, 2015). Similarly to how *concreteness fading* is proposed as a way to improve transfer by resorting to increasingly abstract examples (Fyfe, McNeil, Son, & Goldstone, 2014), it may be a promising route to develop a *semantic congruence fading* process using increasingly incongruent examples. In the case of the problems used in the current study, starting with teaching the 1-step algorithm on ordinal problems, then moving to hybrid problems and then to concrete problems may be a way to help learners acquire a better understanding of this algorithm, and consequently learn to use it in any situation, regardless of the semantics conveyed by the problem statement.

As previously mentioned, content effects such as those revealed through the six experiments of this paper do not fall within the scope of the schema theory (Kintsch & Greeno, 1985) nor of the mental model approach (Johnson-Laird, 1983; Reusser, 1990). For instance, Experiment 4's results showed that participants use different solving algorithms to solve isomorphic problems, depending on the type of quantities these problems mention. In the seminal description of the schema theory, Kintsch and Greeno (1985) indicate that the activation of a schema to solve a problem is done based on the propositional structure of the problem statement. In other words, the relevant entities are the numerical values and the relations explicitly described in the problems. The model proposed by Kintsch and Greeno does not explicitly cover the possibility that the world knowledge evoked by the elements mentioned in a problem may influence solvers' representation of the problems, nor their choice of a solving algorithm. Similarly, the mental model theory – as operationalized in Staub and Reusser's (1995) Situation-Problem-Solver model – carries the idea that a situation model necessarily corresponds to the structure of the depicted situation. Thus, this approach struggles to account for the idea that the initial problem representation constructed by the solvers may vary depending on the state of their general, non-mathematical knowledge about the elements present in the problems. Besides, the notion that participants may overcome their initial encoding of a problem and construct a new representation also falls outside of the Situation-Problem-Solver's scope. In Experiments 5 and 6, participants were shown to engage in a semantic recoding process, akin to re-representation (Vicente, Orrantia, & Verschaffel, 2007), when their initial representation of the problem led to a dead end. We believe that the existence of these alternative problem encodings goes beyond what the Situation-

Problem-Solver may be able to predict. The idea that an interpreted structure could be recoded into a new representation when necessary seems to be a crucial issue deserving attention for upcoming arithmetic word problem solving frameworks.

Altogether, we believe that the results of the six experiments showing the influence of cardinality and ordinality on the encoding, recoding and solving of mathematical word problems call for the creation of a model of mathematical word problem solving encompassing the influence of our daily-life knowledge as well as the central role of semantic recoding into a more accurate depiction of the interpretative processes at play in mathematical reasoning.

References

- 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.
- Bassok, M. (2001). Semantic alignments in mathematical word problems. *The analogical mind: Perspectives from cognitive science*, 401-433.
- Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology*, 35, 99-134.
- Bassok, M., Pedigo, S. F., & Oskarsson, A. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 343–352.
- Bassok, M., Wu, L. L., & Olseth, K. L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory and Cognition*, 23, 354-367.
- Bermejo, V. (1996). *Cardinality development and counting*. *Developmental Psychology*, 32(2), 263-268.
- Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3), 792.
- Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: evidence for a mental time line. *Neuroscience & Biobehavioral Reviews*, 36(10), 2257-2273.
- Boroditsky, L. (2011). How language shapes thought. *Scientific American*, 304(2), 62-65.
- Carpenter, T. P., & Moser, J. M. (1982). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective*, (pp. 9-24). Hillsdale, NJ: Erlbaum.
- Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language learning*, 58, 63-79.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.

- Clément, E., & Richard, J.-F. (1997). Knowledge of domain effects in problem representation: The case of Tower of Hanoi isomorphs. *Thinking & Reasoning*, 3(2), 133-157.
- Colomé, À., & 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.
- Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, 137(1), 22-38.
- Coquin-Viennot, D., & Moreau, S. (2003). Highlighting the role of the episodic situation model in the solving of arithmetical problems. *European Journal of Psychology of Education*, 3, 267-279.
- Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive psychology*, 20(4), 405-438.
- Dantzig, T. (1945). *Number the language of science. A critical survey written for the cultured non-mathematician*. New York, NY: 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.
- Davis, P., Hersh, R., & Marchisotto, E. A. (2011). *The mathematical experience*. Springer Science & Business Media.
- Day, S. B., Motz, B. A., & Goldstone, R. L. (2015). The cognitive costs of context: The effects of concreteness and immersiveness in instructional examples. *Frontiers in Psychology*, 6, 1876.
- 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.
- Delazer, M., & Butterworth, B. (1997). A dissociation of number meanings. *Cognitive Neuropsychology*, 14(4), 613-636.
- Devidal, M., Fayol, M., & Barrouillet, P. (1997). Stratégies de lecture et résolution de problèmes arithmétiques. *L'Année Psychologique*, 97(1), 9-31.
- Droit-Volet, S., & Coull, J. (2015). The developmental emergence of the mental timeline: spatial and numerical distortion of time judgement. *PloS one*, 10(7), e0130465.

- Fayol, M., Abdi, H., & Gombert, J. E. (1987). Arithmetic problems formulation and working memory load. *Cognition and Instruction*, 4(3), 187-202.
- Fischbein, H. (1987). *Intuition in science and mathematics: An educational approach*. (Vol. 5). Springer Science & Business Media.
- Fischer, F. E., & Beckey, R. D. (1990). Beginning kindergarteners' perception of number. *Perceptual and Motor Skills*, 70(2), 419-425.
- Frege, G. (1980). *The Foundations of Arithmetic* [1884], transl. J. Austin. Evanston, IL: Northwestern University Press.
- Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, 34(8), 1430-1451.
- Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.
- Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, 26(1), 9-25.
- Gamo, S., Sander, E., & Richard, J.-F. (2010). Transfer of strategy use by semantic recoding in arithmetic problem solving. *Learning and Instruction*, 20, 400-410.
- Gelman, R., & Gallistel, C. R. (1978). *The child's concept of number*. Cambridge, MA: Harvard University Press.
- Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 65(2-3), 263-297.
- Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52(2), 125-157.
- Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive psychology*, 46(4), 414-466.
- Greer, B. (1992). Multiplication and division as models of solutions. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 276-295). New York: Macmillan.
- 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*, online first.
- Hofstadter, D. R., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. New York: Basic Books.

- Hood, B. M. (1998). Gravity does rule for falling events. *Developmental Science*, 1(1), 59-63.
- Hood, B. M., Santos, L., & Fieselman, S. (2000). Two-year-olds' naïve predictions for horizontal trajectories. *Developmental Science*, 3(3), 328-332.
- Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. *Child Development*, Vol. 54, No. 1, 84-90.
- Izard, V., Streri, A., & Spelke, E. S. (2014). Toward exact number: Young children use one-to-one correspondence to measure set identity but not numerical equality. *Cognitive Psychology*, 72, 27-53.
- Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, 4(1), 71-115.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Harvard University Press.
- Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, 107(43), 18243-18250.
- Kim, I. K., & Spelke, E. S. (1992). Infants' sensitivity to effects of gravity on visible object motion. *Journal of Experimental Psychology: Human Perception and Performance*, 18(2), 385.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review*, 95(2), 163.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological review*, 92(1), 109.
- Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being* (Nachdr.). New York, NY: Basic Books.
- 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.
- Lee, H. S., DeWolf, M., Bassok, M., & Holyoak, K. J. (2016). Conceptual and procedural distinctions between fractions and decimals: A cross-national comparison. *Cognition*, 147, 57-69.
- Meyer, C., Barbiers, S., & Weerman, F. (2016). Order and Ordinality: The Acquisition of Cardinals and Ordinals in Dutch. In Boston *University Conference on Language Development* (BUCLD) 40. Cascadilla.

- 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.
- 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.
- Nesher, P., Greeno, J. G., & Riley, M. S. (1982). The development of semantic categories for addition and subtraction. *Educational Studies in Mathematics*, 13(4), 373-394.
- Northcote, M. T., & Marshall, L. (2016). What mathematics calculations do adults do in their everyday lives?: Part 1 of a report on the Everyday Mathematics Project. *Australian Primary Mathematics Classroom*, 21 (2), 8-17.
- Piaget, J. (1965). *The child's conception of number*. New York: W. W. Norton. (Original work published 1941).
- 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*, Vol. II. New York: Pergamon Press.
- Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). *Teaching the conceptual structure of mathematics*. *Educational Psychologist*, 47(3), 189-203.
- Riley, M. S., Greeno, J. G., & Heller, J. (1983). Development of children's problem-solving ability in arithmetic. *The development of mathematical thinking*, 153-196.
- Rivet, A. E., & Krajcik, J. S. (2008). Contextualizing instruction: Leveraging students' prior knowledge and experiences to foster understanding of middle school science. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 45(1), 79-100.
- Rubinsten, O., Dana, S., Lavro, D., & Berger, A. (2013). Processing ordinality and quantity: ERP evidence of separate mechanisms. *Brain and Cognition*, 82(2), 201-212.
- Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Erlbaum.

- Russell, B. (1903). *Principles of mathematics*. Cambridge, UK: Cambridge University Press.
- Russell, B. (1919). *Introduction to mathematical philosophy*. London: Allen & Unwin.
- Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, 108(3), 662-674.
- Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology*, 103(3), 325-337.
- Schank, R. C. (1975). The role of memory in language processing. In: C. N. Cofer (ed), *The Structure of Human Memory*. San Francisco. (1975): p. 162-189
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans and understanding: An inquiry into human knowledge structures*. New York: Hillsdale.
- Schoenfeld, A. (2009). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. *Colección Digital Eudoxus*, (7).
- Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A “non-numerical” account. *Cognitive Development*, 12(3), 349-372.
- Son, J. Y., & Goldstone, R. L. (2009). Contextualization in perspective. *Cognition and Instruction*, 27(1), 51-89.
- Squire, S., & Bryant, P. (2002). From sharing to dividing: Young children’s understanding of division. *Developmental Science*, 5(4), 452-466.
- Stacey, K. (2005). The place of problem solving in contemporary mathematics curriculum documents. *The Journal of Mathematical Behavior*, 24(3-4), 341-350.
- Staub, F. C., & Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension* (pp. 285-305). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Suppes, P. (1972). *Axiomatic set theory*. New York, NY: Dover Publications.
- Suzuki, R., & Shimodaira, H. (2006). Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), 1540-1542.
- Thevenot, C. (2010). Arithmetic word problem solving: Evidence for the construction of a mental model. *Acta psychologica*, 133(1), 90-95.
- Thevenot, C., Devidal, M., Barrouillet, P., & Fayol, M. (2007). Why does placing the question before an arithmetic word problem improve performance? A situation model account. *The Quarterly Journal of Experimental Psychology*, 60(1), 43-56.

- Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representations in arithmetic word problem solving. *Quarterly Journal of Experimental Psychology-A*, 58(7), 1311-1323.
- Trench, M., & Minervino, R. A. (2015). The role of surface similarity in analogical retrieval: Bridging the gap between the naturalistic and the experimental traditions. *Cognitive science*, 39(6), 1292-1319.
- Turconi, E., Jemel, B., Rossion, B., & Seron, X. (2004). Electrophysiological evidence for differential processing of numerical quantity and order in humans. *Cognitive Brain Research*, 21(1), 22-38.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Vergnaud, G. (1983). Multiplicative structures. *Acquisition of mathematics concepts and processes*, 127-174.
- Vicente, S., Orrantia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology*, 77(4), 829-848.
- 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.
- Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin & Review*, 15(2), 426-430.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358(6389), 749-750.

CHAPTER 3 - TROUBLE DOWN MEMORY LANE: MISREMEMBRANCE REVEALS ENCODING DIFFERENCES IN ARITHMETIC WORD PROBLEMS



Presentation

Chapter 2 made a case for the influence of cardinal and ordinal quantities on the interpretation of arithmetic word problems. Building upon this work, Chapter 3 uses similar materials to go one step further and investigate the nature of the representations constructed in working memory while solving cardinal and ordinal problems. In particular, the recollection of problem statements is investigated, in parallel to the strategies developed to solve them.

Resting upon the assumption that an individual's representation of a problem is reflected in their recollection of its statement, this work investigates what makes ordinal representations so different from cardinal ones. Based on the hypothesized structure of ordinal encodings, we predict that participants will include a specific inference in their representation of these problems, that they will not include in a cardinal representation. We investigate whether participants erroneously recall or recognize this inference more often on one of the two types of problems.

Experiment 1 is a recall task in which participants had to solve two problems before being presented with an unexpected recall task in which they had to write down the problem statements as accurately as possible. Experiment 2 uses a similar design, but with double the number of problem statements, and a cued recall paradigm. Experiment 3 presents participants with 18 problems to solve, and then unexpectedly asks them to identify experimenter-induced changes in the problems. In all three experiments, we predict that on ordinal problems, participants will tend to erroneously recall or recognize a sentence describing a piece of information that was not present in the problem statements, but that could be inferred from an ordinal representation of the situation. Recall mistakes between cardinal and ordinal problems are compared, and the link between participants' algorithm use and their propensity to make such mistakes is investigated.

By probing participants' memory of the problem statements, this chapter offers converging evidence for the influence of cardinal and ordinal quantities on problem representation. Additionally, the use of problems written in French in Experiments 1 and 2, and in English in Experiment 3 provides some indication as to the cross-linguistic robustness of the effects demonstrated in this thesis.

Abstract

Is there a fundamental difference between counting years and kilograms? Marbles and centimeters? Floors and euros? Recent evidence suggests that non-mathematical world knowledge unrelated to the mathematical structure of a problem can nevertheless influence its semantic encoding. To tackle this question, we created arithmetic word problems devised to promote contrasting encodings by featuring different quantities, in French and in English. We designed three experiments investigating the representations constructed and memorized by 302 adult participants when solving the problems. After an initial solving task, participants were given an unexpected task: either recall the problems (Experiments 1 and 2) or identify experimenter-induced changes in target problem sentences (Experiment 3). We predicted that the use of specific quantities in the problem statements was enough to lead participants to erroneously recall mathematical information that was not present in the problems, but that could be inferred from one of the two possible encodings of the situations. Results across all three experiments consistently indicate that participants construct and memorize a different problem encoding depending on the quantities involved. They misremembered problems involving durations, heights, or elevators by including new information into their problem representation. The same recall mistakes were not made for problems involving prices, weights or collections. This supports the claim that knowledge related to daily-life quantities substantially influences arithmetic reasoning, despite such knowledge being irrelevant for abstract reasoning.

Keywords

sentence recognition · semantic encoding · mental models · arithmetic problems

Introduction

A recurring question in the arithmetic problem solving literature regards the nature of the problem representations constructed by the solvers in working memory. Be it through the implementation of problem schemata akin to behavioral scripts (Kintsch & Greeno, 1985; Schank & Abelson, 1977), the construction of mental models depicting the problem situation (Johnson-Laird, 1983; Staub & Reusser, 1995), or the abstraction of an interpreted structure describing the solvers' interpretation of a given problem statement (Bassok, 2001), different theories have attempted to model the representational aspects of arithmetic word problem solving. In this paper, we assess the validity of an emerging hypothesis regarding the representation of numerical situations by simultaneously gathering evidence from three distinct measures of problem representations: strategy choice, text recall and sentence recognition.

The hypothesis that we intend to investigate comes from prior work conducted by Sander and his colleagues, who have argued that the representations of numerical situations, especially in arithmetic word problems, tend to fall within one of two categories: cardinal encodings or ordinal encodings (Gamo, Sander, & Richard, 2010; Gros, Sander, & Thibaut, 2016, 2019; Gros, Thibaut, & Sander, 2017). The notion of ordinality and the notion of cardinality express two sides of numbers: their existence as an item in an ordered list, and their meaning as the total number of entities being counted. This distinction is fundamental in mathematics (Dantzig, 1945; Frege, 1980; Russell, 1919), especially in set theory (Dauben, 1990; Suppes, 1972), and several works in developmental psychology have shown that it has implications reaching beyond the realm of formal mathematics.

Indeed, in the seminal research of Gelman and Gallistel (1986) on the development of counting in children, it was proposed that five counting principles need to be mastered by children in their efforts to become proficient counters. Among those, the “stable-order principle” refers to the development of ordinality, that is, knowing that the list of words used to count needs to be said in a definite and stable order, each word having the same predecessor and the same successor over trials. The “cardinal principle”, on the other hand, refers to the understanding that the final word of an enumeration indicates the total number of entities in the set being counted, thus being an indicator of the development of the notion of cardinality. Following this work, the development of the cardinal meaning of numbers in the early years of

life has been the focus of an important number of studies with children (e.g. Bermejo, 1996; Le Corre & Carey, 2007; Sarnecka & Lee, 2009; Wynn, 1992) and the development of the ordinal use of numbers by pre-school children has also been under scrutiny in a few experimental works (Fischer & Beckey, 1990; Miller, Major, Shu, & Zhang, 2000; Miller, Marcovitch, Boseovski, & Lewkowicz, 2015). In recent years, the differences in the development of these two sides of counting procedures has been under direct investigation, thus showing the growing interest for this distinction in developmental psychology (Colomé & Noël, 2012; Meyer, Barbiers, & Weerman, 2016; Wasner, Moeller, Fischer, & Nuerk, 2015). However, despite the importance of this literature on the first steps of learning how to count, few works have been conducted to investigate the role that cardinality and ordinality still hold in adults, when conceiving of general numerical situations.

Regarding this question, a growing body of research seems to indicate that even after counting procedures are acquired, there remains an ontological difference between the way we conceive of numbers either as order labels or as count values. As previously mentioned, this idea was born from Gamo et al.'s (2010) work on problems sharing the same mathematical structure but admitting two distinct solving algorithms. They showed that participants' choice of solving algorithms depended on whether the problems featured quantities that could be represented along an ordered axis, such as a timeline (duration problems) or whether they used quantities that we tend to conceive of as unordered count values (number-of-people problems and price problems). Indeed, in their experiment, a timeline representation of the situation made it possible to perceive the relevance of the shortest algorithm to solve the problems, whereas a cardinal representation of the situation only let solvers use the longer algorithm involving more steps (Gamo et al., 2010). This difference was attributed to the idea that a timeline encoding lets solvers make a specific type of inference that can hardly be made with a cardinal encoding. Consider, for instance, the following problem:

Paul has 8 red marbles. He also has blue marbles. In total, Paul has 14 marbles. Jolene has as many blue marbles as Paul, and some green marbles. She has 3 green marbles less than Tom has red marbles. How many marbles does Jolene have?

This problem involves counting marbles, and thus emphasizes the cardinal nature of the numbers it features. There is no reason to mentally line up the marbles in a

specific order, and thus we tend to think of the marbles of different colors as distinct, autonomous entities organized as subsets to be combined. Participants thus tend to consider that the only way to calculate the number of marbles that Jolene has is to calculate the number of blue marbles she has and add it to the number of green marbles she has. That is, most participants use a 3-step algorithm to solve this problem: $14 - 8 = 6$; $8 - 3 = 5$; $6 + 5 = 11$. They identify that Jolene has 6 blue marbles and 5 green marbles, thus adding up to 11 marbles in total. On the other hand, consider the following duration problem:

The construction of the palace took 8 years. Plans for the construction were made beforehand. The construction of the palace was completed in year 14.

The construction of the castle started at the same time as the construction of the palace. The construction of the castle took 3 years less than the construction of the palace. When was the construction of the castle completed?

This problem has the same mathematical structure as the marble problem, but because it involves duration values instead of marble counts, we tend to conceive of the described situation as being ordered along a timeline. The different entities are not represented as parts and wholes, but as states and transitions along an axis (Gamo et al., 2010). This lets us see that there is a much shorter solving algorithm to be found: $14 - 3 = 11$. This algorithm could also have been used to solve the marble problem, but participants rarely manage to do so. Indeed, using the 1-step algorithm on the marble problem would require making the inference that since Tom and Jolene both have the same number of blue marbles, then there is no need to calculate this number, nor to calculate how many green marbles Jolene has. Instead, one needs to infer that since Jolene has 3 green marbles less than Tom has red marbles, and since they both have the same number of blue marbles, then Jolene simply has 3 marbles less than Tom in total. On the duration problem, on the other hand, a timeline representation allows for direct comparison of the time it took to build the palace and the time it took to build the castle. The salience of this comparison makes it easier to infer that since both constructions started at the same time, and since the construction of the castle took 3 years less than the construction of the palace, then the construction of the castle was completed 3 years before the construction of the palace. In other words, in both problems it is unnecessary to calculate the value of the common part to find the solution. Fig. 1 describes the deep mathematical structure shared by these two problems, as well as the representations that are thought to be constructed, resulting

in different solving algorithms being used. In order to use the 1-step algorithm “*Whole 1 – Difference = Whole 2*”, one needs to understand that the difference described between *Part 1* et *Part 3* is equal to the difference between *Whole 1* and *Whole 2*. According to the solving algorithms used by participants, this inference is significantly easier to make on ordinal than on cardinal problems (Gamo et al., 2010).

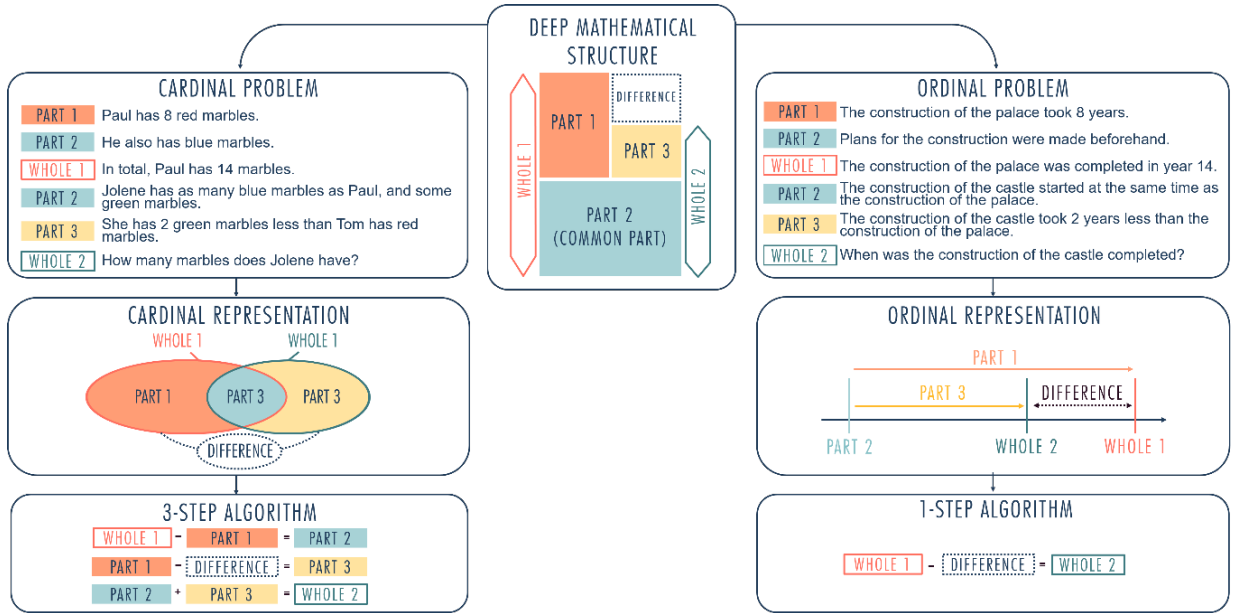


Fig. 1 Graphical summary of the encoding differences between cardinal and ordinal problems.

Subsequent works on problems admitting multiple solving algorithms revealed that different ordinal quantities (durations, height, floors) and different cardinal quantities (collections of elements, price weight) could replicate this effect on algorithm choice (Gros et al., 2016; Gros, Thibaut, & Sander, 2015). Additionally, results from a drawing task with children and adults in Gros et al. (2016) brought converging evidence regarding the cardinal versus ordinal nature of the problems’ representation underlying the algorithm choice. This distinction was even shown to affect expert mathematicians in their assessment of the solvability of cardinal problems (Gros et al., 2019). In this paper, we intend to demonstrate that the difference between cardinal and ordinal quantities is so substantial that it influences the very representations stored in memory and can lead to the formation of false memories by the participants.

Since no means of direct investigation of the representations themselves are available, we believe that text recall tasks and sentence recognition tasks can bring valuable information regarding the nature of the representations built: if there is indeed an ontological difference between the representations we encode of ordinal and cardinal situations, then this difference should result in different encodings being

constructed and memorized. We strove to assess the validity of this claim by evaluating the presence of specific inferences that can only be drawn from the problem text if participants encode an ordinal representation of the situation. Indeed, previous works on text comprehension suggest that sentences from which inferences can be drawn may mislead participants both in recognition (Bransford & Johnson, 1973; Kintsch & Bates, 1977; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Noordman & Vonk, 2015) and in recall tasks (Black & Bern, 1981; Corbett & Doshier, 1978; Kintsch & Van Dijk, 1978; Sulin & Dooling, 1974). In Bower, Black and Turner's (1979) famous work on the importance of scripts in text comprehension, it was shown that participants tended to infer actions that were not explicitly described in the text but that were coherent with the scenario depicted by the text. Those inferences then led participants to erroneously recall events that were never described in the text, with a surprisingly high degree of confidence in their recalls. Here, we predicted that the implicit inferences drawn from one of the two possible encodings of the problems would lead participants to erroneously remember pieces of information which were not initially present in the problems, but which could be inferred from an ordinal representation of the situation.

Our work thus builds on previous paradigms using recognition tasks to evaluate which inferences were included in participants' representation of a given situation. For instance, in one of their experiments, Mani and Johnson-Laird (1982) designed a task to investigate participants' mental representation of spatial descriptions. They presented participants with 4-sentence descriptions of spatial configurations of the form "A is to the left of B. C is to the right of B. D is in front of A. E is in front of B." Participants had to evaluate whether specific diagrams respected the configuration described in the previous statement. After the task, participants were presented with an unexpected recognition task in which they had to identify among a series of 4-sentence statements which were the ones they had been presented before, and which were new statements previously unseen. The authors used three types of test statements: some were identical to the ones previously seen, some described different spatial configurations from the ones previously seen, and some presented a spatial configuration that was inferable from the statements previously seen, although the propositional structure of the text itself was different. Interestingly, the authors showed that the statements presenting an inferable spatial configuration tended to be erroneously recognized more often than the ones presenting a different

spatial configuration. In other words, participants tended to base their recognition of the texts on the mental model they had constructed of the problems, instead of only using the verbatim text as a basis for recognition. This property of text recognition is especially relevant to our study, as it indicates that misremembrance can inform on the representations constructed by the participants.

In a similar spirit, Verschaffel (1994) used a solving task followed by a retelling task to investigate the nature of the representations constructed by fifth graders when solving arithmetic word problems. They found that children tended to reword inconsistent problems and retell them as consistent problems, thus failing the retell task, whereas they did not perform such rewording on already consistent problems. This was interpreted as a clear indicator that children's mental representation of the problems differed from their exact wording. In 1995, Hegarty, Mayer, and Monk made another attempt to investigate the representations constructed by individuals engaged in arithmetic word problems solving. They conducted a problem solving task followed by a text recall task and a text recognition task to evaluate which participants constructed a problem model. They showed that the more successful solvers tended to make a lower number of semantic mistakes but a higher number of literal mistakes than the less successful solvers. The authors interpreted this finding as proof that the more proficient participants had constructed a problem model, since they recalled the semantic structure of the problems successfully, but were less accurate in recalling the exact wording of the problems. On the other hand, the less proficient participants likely did not construct a problem model, since they recalled the problems wording accurately but made more mistakes with regards to their semantic structure. However, it has been argued since then that problems that are failed are unlikely to be recognized (Cummins, Kintsch, Reusser, & Weimer, 1988), which suggests that the answers of the less proficient solvers could not be interpreted as proof that they did not construct a problem model (Thevenot, 2010).

More recently, another study using a text recognition task replicated this finding in the field of arithmetic word problem solving (Thevenot, 2010). The author asked participants to solve a series of problems, and then presented them with an unexpected recognition task. Ingeniously, she used three different types of problems in the recognition task: (i) problems identical to the original problems, (ii) new problems mathematically inconsistent with the original ones despite differing by only one or two words, (iii) paraphrastic problems mathematically consistent with the

original problems despite differing by a total of three words. Results revealed that the paraphrastic problems were more often erroneously recognized than were the new inconsistent problems, despite the latter differing by a lower number of words from the original problems. This indicated that participants had constructed and memorized a representation of the problems that depended on the structure of the situations they described rather than on their precise wording.

Building upon these previous studies on text recognition and text recall, our goal in the present study was to show that the representations constructed while solving arithmetic word problems vary significantly depending on the cardinal versus ordinal nature of the quantities they feature, to the point that performance on problem recall and problem recognition tasks are directly influenced by this distinction. More precisely, we predict that the distinction between problems using cardinal quantities and problems using ordinal quantities will be so consequential to the solvers that it will be reflected in their reminiscence of the problems, leading them to construct false memories based on their interpretation of the situation. In this perspective, we designed three experiments in which we presented participants with a solving task followed by an unexpected task scrutinizing their recollection of the problems. In the first two experiments, the solving task was followed by an unexpected recall task in which participants had to write down the problems they just solved from memory. We expected participants to make specific mistakes in the recall task on ordinal problems but not on cardinal problems, due to additional information being automatically inferred from ordinal representations but not from cardinal representations. In a third experiment, we presented participants with an unexpected sentence recognition task, in which they had to decide whether target sentences were included in the problems they previously solved, or whether these sentences had been modified by the experimenter. We expected participants to be more likely to erroneously recognize modified sentences of ordinal than of cardinal problems, due to the modified version presenting information that might have been automatically inferred from an ordinal representation but not from a cardinal representation. The first two experiments were conducted in French, the third one in English.

Experiment 1

Experiment 1 was a first attempt to gather evidence regarding the inferences that can be drawn from an ordinal representation but not from a cardinal representation. We formulated three hypotheses regarding the solving task and the problem recall task. First, during the solving task, we predicted that we could replicate the results from Gros et al. (2017), meaning that due to the different representations of the problems, participants' ability to use the shortest algorithm would depend upon the cardinal versus ordinal nature of the problems: participants should have an easier time using the 1-step algorithm on duration problems than on collection problems. Second, regarding the unexpected recall task, we formulated a hypothesis pertaining to the type of recall mistakes made by the participants. We assumed that if participants constructed on ordinal representation of the situation, then they would be more likely to make an automatic inference regarding the difference mentioned in the problems. Precisely, while the 5th sentence of the problems always introduced a difference between *Part 3* and *Part 1* (e.g. "Jolene has 2 green marbles less than Tom has red marbles" in a cardinal problem or "The construction of the castle took 2 years less than the construction of the palace" in an ordinal problem) we predicted that an ordinal representation would make it easy to infer that this difference was equal to the difference between *Whole 2* and *Whole 1* whereas the same inference would be harder to make when a cardinal representation was constructed. Participants would thus integrate this inference into their representation of ordinal problems (e.g. "The construction of the castle ended 2 years before the construction of the palace"), but not cardinal problems (e.g. "Jolene has 2 marbles less than Tom"). This piece of knowledge is not directly present in the problem statement, but can be inferred from the pieces of information provided. Thus, we made the hypothesis that participants would tend to erroneously recall a sentence describing a difference between *Whole 2* and *Whole 1* instead of the original sentence describing the difference between *Part 3* and *Part 1*, more often on ordinal than on cardinal problems. Finally, we made the hypothesis that this type of recall mistake would be more likely to occur on problems solved using the 1-step algorithm, denoting an ordinal encoding, than on problems solved using the 3-step algorithm.

Method

Participants. Participants were recruited online through social networks and mailing lists. They participated voluntarily without any monetary incentive. It was decided that the survey would be closed 2 weeks after its online broadcast. Out of the 140 people who participated, 13 left at least one of the questions unanswered and were subsequently excluded from our dataset. The analyses were conducted on the remaining 127 participants (81 women, mean age = 33.39 years, $SD = 8.15$). Participants all spoke French fluently.

Materials. The materials constructed were based on previous works focusing on the difference between cardinal and ordinal encodings (Gamo et al., 2010; Gros et al., 2017). To maximize the encoding difference between the problems, we selected our materials from the two most stereotypical quantities used in Gros et al. (2017): collection problems (see Table 1., column A.) and duration problems (see Table 1., column B.). We used a within-subject design to allow for within-subject comparisons between responses on cardinal and on ordinal problems. Each participant was presented with one randomly chosen cardinal problem (a collection problem) and one randomly chosen ordinal problem (a duration problem). All the problems were isomorphs and the numerical values used were randomized across problems, according to the following rule: $15 \geq \textit{Whole} > \textit{Part} > 4 > \textit{Difference} \geq 2$.

Table 1 Cardinal and ordinal problem statements used in Experiment 1.

A. Cardinal problems	B. Ordinal problems
Paul has 5 red marbles. He also has blue marbles. In total, Paul has 14 marbles. Jolene has as many blue marbles as Paul, and some green marbles. She has 2 green marbles less than Tom has red marbles. How many marbles does Jolene have?	Sofia travelled 5 hours. Her trip started during the day. Sofia arrived at 14 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?
Sarah owns 5 goldfish. Her other pets are all iguanas. In total, she owns 14 pets. Bobby is pet-sitting Sarah's iguanas during the holidays, he puts them with his pet turtles. Bobby owns 2 turtles less than Sarah owns goldfish. How many pets are there at Bobby's?	The construction of the palace took 5 years. Plans for the construction were made beforehand. The construction of the palace was completed in year 14. The construction of the castle started at the same time as the construction of the palace. The construction of the castle took 2 years less than the construction of the palace. When was the construction of the castle completed?

Procedure. This experiment was conducted online using the Qualtrics platform for online experiments. On the first page, the instructions read:

On the next page, you will find an arithmetic problem. Please take the time to read it carefully. Your task is to try to solve the problem using as few operations as possible. We ask that you take enough time to read and understand the problem, as this is not a speed test. Remember that the goal is to solve the problems using as few operations as possible. Type down every operation(s) that you used to come up with the solution, even the simplest one(s) that you can mentally calculate. For instance, the computation “ $15 - 6 - 2 = 7$ ”, should not be written as a unique operation, but broken down as “ $15 - 6 = 9$ ” and “ $9 - 2 = 7$ ”, which then count for two operations. (translated from French).

On the next page, a problem was presented, either evoking a cardinal encoding (collection problem) or an ordinal encoding (duration problem). Along with the problem statement, the instructions to solve the problem using as few operations as possible were reminded to the participants. Below the problem a text box allowed participants to write down the operation(s) they used, and another text box was used to write down the problem’s solution. On the next page, the initial instructions were repeated, and another problem followed. Depending on which problem participants had been presented initially, the second problem was chosen to evoke a different encoding. In other words, if participants had to solve a cardinal problem first, then the second problem was ordinal. Problem order was randomized between participants. When participants had solved both problems, they were presented with an unexpected recall task. They were told that they had to recall as precisely as possible the text from the first of the two problems. They were instructed to write everything they remembered about the problem statement as faithfully as possible. After they had completed this task, the next page asked them to write down the text of the second problem they had to solve. In average, participants completed the experiment in 19 minutes and 26 seconds.

Results

First, we analyzed participants’ answers to the solving task. In 95.21% of the cases, the algorithms used by the participants to solve the problems could easily be inferred from their report of the operations they used to solve the problems. The 4.79% of cases where the algorithm could not be directly inferred from their response (correct

response provided with no operation leading to it) were classified as “unidentified” (see Fig. 2).

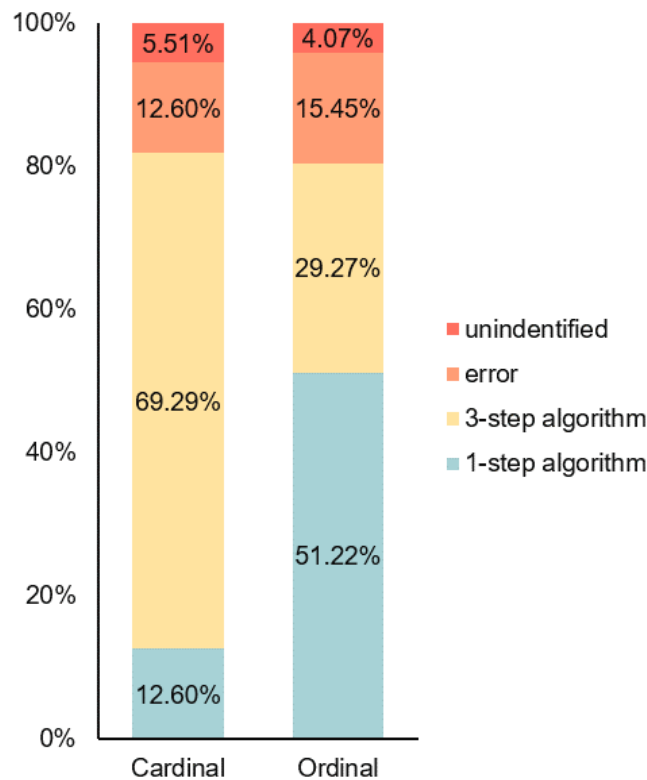


Fig. 2 Strategy distribution depending on the quantities used in the problems.

The identifiable responses were either classified as “1-step algorithm” (successful use of the shortest algorithm), “3-step algorithm” (successful use of the longest algorithm) or “error” (wrong operations leading to a false answer). The distribution of the participants’ solving strategies depending on the ordinal versus cardinal nature of the problems is described in Fig. 2. We used a generalized linear mixed model (GLMM) with a binomial distribution to evaluate how likely participants were to use the 1-step algorithm on cardinal and on ordinal problems. We chose the successful use of the 1-step algorithm as the dependent variable and the semantic nature (cardinal versus ordinal) of the problems as a fixed effect. We accounted for each participant solving each type of problem by including a random effect for each respondent. We performed the analyses using R (R Core Team, 2019) and lme4 (Bates, Maechler, Bolker, & Walker, 2015). The model successfully converged, with a total explanatory power of 49% (conditional R^2). As in Gamo et al. (2010), and Gros et al. (2017), the participants were considerably more likely to discover the 1-step algorithm on ordinal problems (51.22%) than they were on cardinal problems (12.60%); $z = 4.67$, $p < .001$.

Second, in order to investigate the nature of the representations encoded by the participants on each type of problem, we analyzed the problem statements they tried to recall. Namely, we studied whether participants had misremembered the 5th sentence of the problems describing the difference between *Part 1* and *Part 3* (see Fig. 1). For each problem recalled, we noted whether the participants had erroneously recalled a sentence describing the difference between *Part 1* and *Part 3* as a difference between *Whole 1* and *Whole 2* instead (a whole-to-whole inference recall mistake). In other words, for the marble problem, we evaluated how often participants recalled “Jolene has x marbles less than Tom” (part-to-part difference) instead of the correct sentence “Jolene has x green marbles less than Tom *has red marbles*” (whole-to-whole difference). For the duration problem presented before, we evaluated how often participants recalled a sentence stating “The construction of the castle ended 3 years before that of the palace” instead of the correct sentence “the construction of the castle took 3 years less than that of the palace.”

Table 2 Distribution of participants’ erroneous whole-to-whole recalls on cardinal and ordinal problems.

		Ordinal problems	
		No whole-to-whole difference recalled	Whole-to-whole difference erroneously recalled
Cardinal problems	No whole-to-whole difference recalled	106	20
	Whole-to-whole difference erroneously recalled	1	0

Table 2 presents the contingency table indicating the distribution of participants’ whole-to-whole recall mistakes on cardinal and on ordinal problems. Participants recalled this inference instead of the proper problem phrasing of the problems in 15.78% of the ordinal problems, whereas they only made this mistake in 0.79% of the cardinal problems. Due to the extremely low number of participants (1 out of 127) who erroneously recalled a sentence indicating a whole-to-whole difference on a cardinal problem, the comparison between cardinal and ordinal problems could not be done using variance analysis. Instead, we performed a McNemar test between these two conditions ($\chi^2(1, N = 127) = 15.43, p < .001$), which indicated that the semantic nature of the problems (ordinal versus cardinal) had a significant impact on the rate of whole-to-whole sentences being erroneously recalled by participants, as hypothesized.

Finally, we investigated which algorithms used in the solving task were the most likely to lead participants to make the erroneous whole-to-whole inference in the recall task. Interestingly, the only occurrence in which a participant had used the 3-step algorithm prior to making the erroneous whole-to-whole inference was also the only occurrence where a participant made this recall mistake on a cardinal problem. The other cases in which a participant erroneously recalled a sentence evoking the whole-to-whole inference mostly regarded problems solved using the 1-step algorithm (71.43% of the recall mistakes), although a small portion of the whole-to-whole recall mistakes were attributable either to answers that were insufficiently detailed to be interpreted (unidentified answers: 9.52%) or to errors in the solving task (14.29%) (see Table 3).

Table 3 Strategies leading to the erroneous recall of whole-to-whole sentences

	No whole-to-whole difference recalled	Whole-to-whole difference erroneously recalled
1-step algorithm	64	15
3-step algorithm	123	1
Error	32	3
Unidentified answer	14	2

Discussion

This experiment provides insights into the problem representations constructed by the participants. First, it replicated and extended Gamo et al.'s (2010) finding by showing that the choice of a solving algorithm was directly dependent on the cardinal versus ordinal nature of the quantities used in the problem; which was a clear indicator that different problem representations had been constructed. Participants had been explicitly instructed to use the shortest algorithm they could think of, using as few operations as possible, but only 12.60% of them managed to find the 1-step algorithm on cardinal problems, whereas more than half of them used the 1-step algorithm to solve the ordinal problems.

Second, the analysis of the recall mistakes made by the participants in the recall task provided new insights regarding the nature of the representations constructed by the participants. As hypothesized, participants were more likely to misremember the sentence describing the difference between *Part 1* and *Part 3* as a difference between *Whole 1* and *Whole 2* on ordinal problems (20 participants) than on cardinal problems (only one participant). Despite being mathematically valid, this piece of information regarding the difference between *Whole 1* and *Whole 2* was not directly present in the original problem statements; participants had to infer it from the situation

described. The fact that participants were significantly more likely to assume that the difference was presented between *Whole 1* and *Whole 2* on the original ordinal problems indicates that this inference was automatically made and was included in their representation of the situation. In other words, it supports the hypothesis that ordinal axis-based representations let the solvers understand that if two events start at the same time and one is x years shorter than the other, then it follows that one ends x years before the other (see Fig. 1). This inference is easy to make when the problem's values are ordered along the same oriented axis. On the other hand, this inference is much harder to make on cardinal representations, since it requires understanding that the cardinality difference between the differing parts of two overlapping set is equal to the cardinality difference between the two sets (see Fig. 1).

Third, the vast majority of participants who erroneously recalled the difference as a difference between wholes instead of a difference between parts did so after solving the problems using the 1-step algorithm. Indeed, none of the participants who had solved the ordinal problem using the 3-step algorithm did this mistake. This corroborates the assumption that both the algorithm choice and the misremembrance of the fifth sentence are converging indicators of the nature of the representations constructed by the participants. Overall, this experiment showed that the differences between the encoding of cardinal and ordinal problems were so potent that they tampered with participants' recollection of the problem statements. Participants falsely remembered sentences that were not present in the problems, due to the altered representation they had constructed. However, the task presented to the participants was relatively easy, and few recall mistakes were made in the recollection task. In an attempt to increase the task difficulty while assessing the replicability of the differences observed, we designed a second experiment in which we doubled the number of problems to solve and recall.

Experiment 2

This second experiment attempted to replicate the first experiment's findings in a slightly different setting, using a higher number of problems to increase task difficulty. The hypotheses were the same as in Experiment 1, since the goal was to identify whether participants would still recall a greater number of whole-to-whole inferences on ordinal than on cardinal problems.

Method

Participants. Participants were students from a second-year university psychology class at the University of Bourgogne. They participated in exchange for course credit. A total of 104 students participated in the experiment. One participant was excluded from the analysis due to all of their answers being insufficiently detailed to be interpreted. The analyses were conducted on the remaining 103 participants (72 women, mean age = 20.43 years, $SD = 1.43$). Participants all spoke French fluently.

Materials. The problems used in this experiment were the same as those used in Experiment 1, the difference being that every participant was asked to solve the 4 problems, instead of 2 problems being randomly selected.

Procedure. The experiment was conducted collectively, in a university classroom. Each participant was given a 5-page booklet with the following instructions written on the front page:

You will find an arithmetic problem on each page of this booklet. Your task is to solve the problems using as few operations as possible. You can use the ‘draft’ area, but please copy in the ‘response’ area all the operations that you used to come up with the solution. We ask that you take enough time to read and understand each of these problems, as this is not a speed test. Remember that the goal is to solve the problems using as few operations as possible. For every problem, we ask that you write down every operation(s) that you used to come up with the solution, even the simplest one(s) that you can mentally calculate. For instance, the computation “ $15 - 6 - 2 = 7$ ”, should not be written as a unique operation, but broken down as “ $15 - 6 = 9$ ” and “ $9 - 2 = 7$ ”, which then count for two operations. (translated from French).

The four following booklet pages were divided in three parts: the problem statement, the “draft” area and the “response” area. Problem order was randomized across booklets, and so were the numerical values used in the problems. When participants were done solving the problems, their booklets were collected, and new booklets were, then, handed out to them. Here, the instructions read “On the following pages, you will be asked to recall the text of the problems you just solved. Try to write down the problem statements as faithfully as possible, from memory.” (translated from French). Then, on each following page the recall of a specific problem statement was cued using a sentence describing the theme of the problem. For example: “Try to write down, from memory, the text of the 1st problem you had to solve, about Tom

and Jolene's marbles." (translated from French). The first problem had to be recalled first, and the second problem last, in order to avoid any recency effect. The first and second sets of booklets were matched so that the recall order was identical to the order in which the problems had been presented in the solving task.

Results

As in Experiment 1, we studied participants' response both in the solving task and in the problem recall task, to investigate the differences between their representations of cardinal and ordinal problems. In 98.79% of the cases, the algorithms used by the participants to solve the problems could easily be inferred from their report of the operations they used to solve the problems. Their responses were either classified as "1-step algorithm" (successful use of the shortest strategy), "3-step algorithm" (successful use of the longest strategy) or "error" (failure to use any relevant algorithm to solve the problem). The 1.21% of answers that were not detailed enough to be analyzed were classified as "unidentified". The distribution of the participants' solving strategies depending on the ordinal versus cardinal nature of the problems is described in Fig. 3. As in Experiment 1, we used a generalized linear mixed model with a binomial distribution to evaluate how likely participants were to use the 1-step algorithm on cardinal and on ordinal problems. The successful use of the 1-step algorithm was the dependent variable and the semantic nature (cardinal versus ordinal) of the problems was a fixed effect. We accounted for each participant solving each type of problem by including a random effect for each respondent. The model successfully converged and accounted for 83% of the total variance (conditional R^2). As expected, the participants were considerably more likely to discover the 1-step algorithm on ordinal problems (60.19%) than they were on cardinal problems (15.53%); $z = 6.49$, $p < .001$.

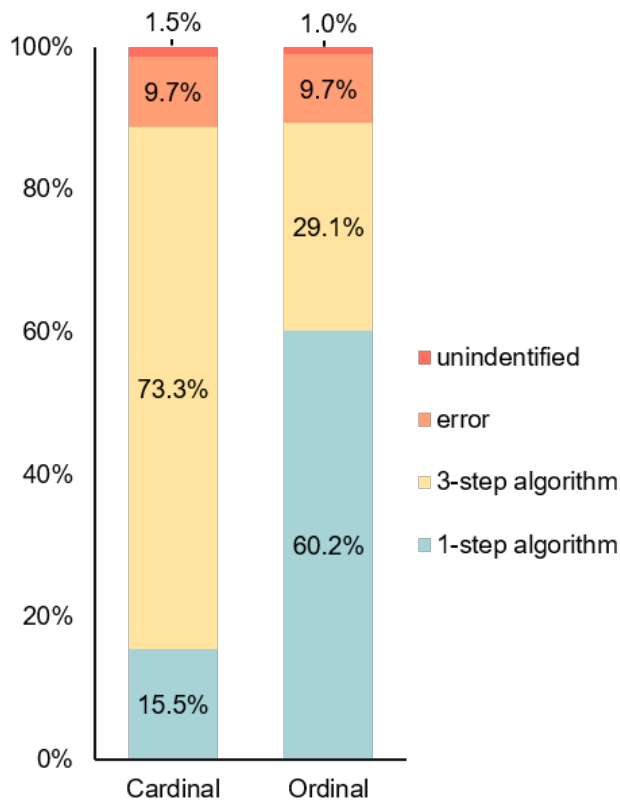


Fig. 3 Strategy distribution depending on the quantities used in the problems.

Table 4 presents the contingency table indicating the distribution of erroneous recalls of a whole-to-whole difference instead of the part-to-part difference on cardinal and on ordinal problems. As in Experiment 1, erroneously recalling a whole-to-whole difference was interpreted as a sign that participants had made the inference that the difference between the parts was equal to the difference between the wholes, which was favored by an ordinal encoding but not by a cardinal encoding. Participants recalled this inference instead of the proper problem phrasing of the problems in 9.71% of the ordinal problems, whereas they never once made this recall mistake on the cardinal problems. Because no participant recalled a whole-to-whole difference instead of the part-to-part difference present in the cardinal problems, we used an exact McNemar test to analyze the difference between the recall mistakes on cardinal and ordinal problems. The test indicated that the semantic nature of the problems (ordinal versus cardinal) had a significant impact on the rate of whole-to-whole sentences being erroneously recalled by the participants ($\chi^2(1, N = 103) = 13.07, p < .001$).

Table 4 Distribution of participants making at least one erroneous whole-to-whole recall on cardinal or ordinal problems

		Ordinal problems	
		No whole-to-whole difference recalled	Whole-to-whole difference erroneously recalled
Cardinal problems	No whole-to-whole difference recalled	88	15
	Whole-to-whole difference erroneously recalled	0	0

Finally, we studied which of the strategies used in the problem solving task were the most likely to lead participants to erroneously recall a whole-to-whole sentence. Since no participant recalled a whole-to-whole sentence on any cardinal problems, all the cases of erroneous inferences came from ordinal problems. Table 5 details the algorithms used prior to misremembering the problems. Interestingly, no use of the 3-step algorithm was ever followed by a whole-to-whole recall mistake. Instead, 90% of the recall mistake came from the use of the 1-step algorithm, 5% were associated with failure to solve the problem in the first place, and 5% were made by participants whose answer on the solving task could not be interpreted for lack of detail.

Table 5 Strategies leading to the erroneous recall of whole-to-whole sentences

	No whole-to-whole difference recalled	Whole-to-whole difference erroneously recalled
1-step algorithm	138	18
3-step algorithm	211	0
Error	39	1
Unidentified answer	4	1

Discussion

This second experiment replicated and extended the findings of Experiment 1. In the solving task, participants remained more prone to use the 1-step algorithm on ordinal than on cardinal problems. In the recall task, the effect observed in Experiment 1 was also replicated. In fact, the distinction between cardinal and ordinal problems was so decisive that out of the 206 attempts to recall a cardinal problem, not a single one led to the erroneous recall of a sentence describing a whole-to-whole difference. The recall of ordinal problems, on the other hand, included this recall mistake about once every 10 trials. This suggests that the differences in the representations encoded by the participants significantly influenced their answers on both tasks in this experiment as well. Finally, the analysis of the solving strategies precluding the erroneous recall

of a whole-to-whole difference supported the hypothesis of different representations being encoded. Indeed, 90% of the recall mistakes followed the use of the shortest algorithm, whereas strictly none of the erroneous recalls were preceded by the use of a 3-step algorithm. Thus, the results supported the link between participants' representations of the numerical situations, their use of a specific solving algorithm and their recall of the texts. In fact, the correlation between their choice of a solving algorithm and their propensity to make a recall mistake shows that the recall mistakes cannot be attributed to a difference in wording between cardinal and ordinal problems. Indeed, ordinal problems were solved using the 3-step algorithm in 29.1% of the cases, but none of the recall mistakes followed the use of this algorithm.

Experiments 1 and 2 relied on participants' tendency to make specific mistakes in their attempts to spontaneously recall the problems. The main benefit of this experimental paradigm is that it allows us to investigate the representations that participants constructed, memorized and freely recalled. There was however one aspect on which Experiment 2 fell short: by doubling the number of problems, we were hoping to increase the task difficulty and thus increase the number of recall mistakes. Yet, participants made fewer recall mistakes in this task. A possible explanation for this is the fact that the experimental setting in this experiment (a university classroom) was arguably more conducive to concentration than the online survey used in Experiment 1. Overall, the task remained relatively easy and the number of participants making a recall mistake on the 5th sentence was relatively low. In order to address exactly how often participants construct a representation including the whole-to-whole inference, we designed a third experiment involving a sentence-recognition task. By asking participants to identify experimenter-induced changes in the fifth sentence of the problem statements, we hope to more directly measure their acceptance of sentences presenting a whole-to-whole inference, thus gaining further insights into the nature of their representations.

Experiment 3

In this third experiment, we used a sentence-recognition paradigm to directly investigate whether participants' representations included the whole-to-whole difference. Instead of recording participants' spontaneous mistakes in a recall task, we gave them target sentences presenting the difference in the problems either as a part-to-part difference or as a whole-to-whole one. For this last experiment, we

recruited English-speaking participants both for practical reasons and to strengthen the cross-linguistic robustness of the effects described in the first two experiments.

Method

Participants. We recruited 80 participants residing in the United States through the Amazon Mechanical Turk website. Due to several participants showing difficulty to formulate basic English sentences in their answers, we removed from the analyses the 10 participants who were not native English speakers. Additionally, we removed 3 participants who successfully solved less than 17% of the problems, thus showing poor attention during the task. The analyses were conducted on the remaining 67 participants (26 women, mean age = 39.18, $SD = 10.85$).

Materials. In this third experiment, problems were written in English. Since the overall rate of erroneous recall was relatively low in the previous two experiments, we used a higher number of problems to increase the task difficulty. The 12 problems created in Gros et al. (2017) were used in this experiment. We added 6 new problems to create a pool of 18 problems to choose from. Each participant was presented with the 18 problems. In order to limit the repetitiveness of the task for the participants, we varied the quantities used in the different problems: the pool of ordinal problems was composed of 3 duration problems, 3 height problems and 3 elevator problems, whereas the pool of cardinal problems was composed of 3 collection problems, 3 price problems and 3 weight problems. Each these quantities had been previously tested and approved in previous experiments investigating the role of the cardinal versus ordinal dimension (Gros et al., 2017, 2019). We used a within-subject design to allow for within-subject comparisons between performance on cardinal and on ordinal problems.

In the recognition task, two types of problem statements were presented to the participants: problem statements identical to the original ones, and problem statements in which one sentence had been slightly modified to present the difference as a whole-to-whole difference instead of the part-to-part difference in the original wording. Both versions of each cardinal problem are presented in Table 6, and both version of each ordinal problem are presented in Table 7.

Table 6 Ordinal problem statements used in the solving task (Original) and in the recognition task (Original + Modified). Changes were systematically introduced in the 5th sentence. The problems' numerical values varied between participants.

Original problem statement	Modified problem statement
<p>Sofia travelled 5 hours. Her trip started during the day. 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?</p>	<p>Sofia travelled 5 hours. Her trip started during the day. Sofia arrived at 11 h. Fred left at the same time as Sofia. Fred arrived 2 hours before Sofia. What time was it when Fred arrived?</p>
<p>The construction of the palace took 5 years. Plans for the construction were made beforehand. The construction of the palace was completed in year 13. The construction of the castle started at the same time as the construction of the palace. The construction of the castle took 2 years less than the construction of the palace. When was the construction of the castle completed?</p>	<p>The construction of the palace took 5 years. Plans for the construction were made beforehand. The construction of the palace was completed in year 13. The construction of the castle started at the same time as the construction of the palace. The construction of the castle was completed 2 years before the palace. When was the construction of the castle completed?</p>
<p>Rose took painting lessons for 5 years. She started at a certain age. Rose stopped attending painting lessons at age 14. Ted started taking painting lessons at the same age as Rose. He attended painting lessons for 3 years less than Rose. How old was Ted when he stopped attending painting lessons?</p>	<p>Rose took painting lessons for 5 years. She started at a certain age. Rose stopped attending painting lessons at age 14. Ted started taking painting lessons at the same age as Rose. He stopped attending the lessons 3 years before Rose. How old was Ted when he stopped attending painting lessons?</p>
<p>Naomi took the elevator and went up 7 floors. She left from the floor where her grandparents live. She arrived to the 15th floor. Her brother David also took the elevator from their grandparents' floor. He went up 3 floors less than Naomi. What floor did David arrive to?</p>	<p>Naomi took the elevator and went up 7 floors. She left from the floor where her grandparents live. She arrived to the 15th floor. Her brother David also took the elevator from their grandparents' floor. He arrived 3 floors below Naomi. What floor did David arrive to?</p>
<p>Katherine took the elevator and went up 7 floors. She left from the floor where the gym is. She arrived to the 13th floor. Yoan also took the elevator from the floor where the gym is. He went up 2 floors less than Katherine. What floor did Yoan arrive to?</p>	<p>Katherine took the elevator and went up 7 floors. She left from the floor where the gym is. She arrived to the 13th floor. Yoan also took the elevator from the floor where the gym is. He arrived 2 floors below Katherine. What floor did Yoan arrive to?</p>

Gloria took the elevator and went up 5 floors.
 She left from the floor where her office is.
 She arrived to the 14th floor.
 Her coworker Larry also took the elevator from their office's floor.
 He went up 2 floors less than Gloria.
 What floor did Larry arrive to?

Slouchy Smurf is 6 centimeters tall.
 He climbs on a smurf table.
 He now attains the height of 11 centimeters.
 Grouchy Smurf climbs on the same table as Slouchy Smurf.
 Grouchy Smurf is 2 centimeters shorter than Slouchy Smurf.
 What height does Grouchy Smurf attain when he climbs on the table?

Obelix's statue is 5 meters tall.
 It is placed on a pedestal.
 Once on the pedestal, it attains 12 meters in height.
 Asterix's statue is placed on the same pedestal as Obelix's.
 Asterix's statue is 3 meters shorter than Obelix's.
 What height does Asterix's statue attain when placed on the pedestal?

The giraffe in the zoo is 6 meters tall.
 It climbs on the biggest rock in the park.
 Once on the rock, it attains 11 meters in height.
 The elephant in the zoo climbs on the same rock as the giraffe.
 The elephant is 3 meters shorter than the giraffe.
 What height does the elephant attain when standing on the rock?

Gloria took the elevator and went up 5 floors.
 She left from the floor where her office is.
 She arrived to the 14th floor.
 Her coworker Larry also took the elevator from their office's floor.
 He arrived 2 floors below Gloria.
 What floor did Larry arrive to?

Slouchy Smurf is 6 centimeters tall.
 He climbs on a smurf table.
 He now attains the height of 11 centimeters.
 Grouchy Smurf climbs on the same table as Slouchy Smurf.
 The height Grouchy Smurf attains on the table is 2 centimeters shorter than that of Slouchy Smurf.
 What height does Grouchy Smurf attain when he climbs on the table?

Obelix's statue is 5 meters tall.
 It is placed on a pedestal.
 Once on the pedestal, it attains 12 meters in height.
 Asterix's statue is placed on the same pedestal as Obelix's.
 The height Asterix's statue attains on the pedestal is 3 meters shorter than that of Obelix's statue.
 What height does Asterix's statue attain when placed on the pedestal?

The giraffe in the zoo is 6 meters tall.
 It climbs on the biggest rock in the park.
 Once on the rock, it attains 11 meters in height.
 The elephant in the zoo climbs on the same rock as the giraffe.
 The height he attains on the rock is 3 meters shorter than the giraffe's.
 What height does the elephant attain when standing on the rock?

Table 7 Cardinal problem statements used in the solving task (Original) and in the recognition task (Original + Modified). Changes were systematically introduced in the 5th sentence. The problems' numerical values varied between participants.

Original problem statement	Modified problem statement
<p>Paul has 7 red marbles. He also has blue marbles. In total, Paul has 13 marbles. Jolene has as many blue marbles as Paul, and some green marbles. She has 2 green marbles less than Tom has red marbles. How many marbles does Jolene have?</p>	<p>Paul has 7 red marbles. He also has blue marbles. In total, Paul has 13 marbles. Jolene has as many blue marbles as Paul, and some green marbles. She has 2 marbles less than Tom. How many marbles does Jolene have?</p>
<p>Sarah owns 7 goldfish. Her other pets are all iguanas. In total, she owns 15 pets. Bobby is pet-sitting Sarah's iguanas during the holidays, he puts them with his pet turtles. Bobby owns 3 turtles less than Sarah owns goldfish. How many pets are there at Bobby's?</p>	<p>Sarah owns 7 goldfish. Her other pets are all iguanas. In total, she owns 15 pets. Bobby is pet-sitting Sarah's iguanas during the holidays, he puts them with his pet turtles. Now at Bobby's, there are 3 pets less than there were at Sarah's before the holidays. How many pets are there at Bobby's?</p>
<p>Karl picked 6 tulips. He puts them with the daffodils he gathered. In total, Karl has 11 flowers in his bouquet. In her bouquet, Clarice has as many daffodils as Karl, and some roses. She has 2 roses less than Karl has tulips. How many flowers does Clarice have in her bouquet?</p>	<p>Karl picked 6 tulips. He puts them with the daffodils he gathered. In total, Karl has 11 flowers in his bouquet. In her bouquet, Clarice has as many daffodils as Karl, and some roses. Clarice has 2 flowers less than Karl does. How many flowers does Clarice have in her bouquet?</p>
<p>In the store, Anthony wants to buy a 5-dollar ruler. He also wants a notebook. In total, that will cost him 13 dollars. Julie wants to buy the same notebook as Anthony, and an eraser. The eraser costs 2 dollars less than the ruler. How much will Julie have to pay?</p>	<p>In the store, Anthony wants to buy a 5-dollar ruler. He also wants a notebook. In total, that will cost him 13 dollars. Julie wants to buy the same notebook as Anthony, and an eraser. In total, she will pay 2 dollars less than Anthony will. How much will Julie have to pay?</p>
<p>The first meal on the menu includes a chocolate cake costing 5 dollars. The meal also includes a mushroom omelet. In total, that makes for a 14-dollar meal. The second meal on the menu includes the same mushroom omelet, and an apple pie. The apple pie costs 3 dollars less than the chocolate cake. How much does the second meal cost?</p>	<p>The first meal on the menu includes a chocolate cake costing 5 dollars. The meal also includes a mushroom omelet. In total, that makes for a 14-dollar meal. The second meal on the menu includes the same mushroom omelet, and an apple pie. The second meal on the menu is 3 dollars cheaper than the first meal. How much does the second meal cost?</p>

Tyler wants to buy French fries that cost 5 dollars.
 He will also take a cheeseburger.
 In total, that will cost him 14 dollars.
 Zoey orders a cheeseburger as well, and a milkshake.
 The milkshake costs 2 dollars less than the French fries.
 How much will Zoey pay for her order?

Joe takes a Russian dictionary weighing 5 kilograms.
 He also takes a Spanish dictionary.
 In total, he is carrying 12 kilograms of books.
 Lola takes Joe's Spanish dictionary and a German dictionary.
 The German dictionary weighs 3 kilograms less than the Russian dictionary.
 How many kilograms is Lola carrying?

A bag of pears weighs 6 kilograms.
 It is weighed together with cheese.
 In total, the weighing scale indicates 11 kilograms.
 The same cheese is then weighed together with a milk carton.
 The milk carton weighs 3 kilograms less than the bag of pears.
 What is indicated on the weighing scale now?

On moving day, Ryan is carrying his microwave oven, which weighs 5 kilograms.
 He is carrying his coffee machine at the same time.
 In total, he is carrying 11 kilograms of appliances.
 Felicia takes Ryan's coffee machine from him while carrying a blender.
 The blender weighs 2 kilograms less than the microwave oven.
 How many kilograms of appliances is Felicia carrying?

Tyler wants to buy French fries that cost 5 dollars.
 He will also take a cheeseburger.
 In total, that will cost him 14 dollars.
 Zoey orders a cheeseburger as well, and a milkshake.
 Her order will cost 2 dollars less than Tyler's.
 How much will Zoey pay for her order?

Joe takes a Russian dictionary weighing 5 kilograms.
 He also takes a Spanish dictionary.
 In total, he is carrying 12 kilograms of books.
 Lola takes Joe's Spanish dictionary and a German dictionary.
 In total, Lola's books weigh 3 kilograms less than Joe's.
 How many kilograms is Lola carrying?

A bag of pears weighs 6 kilograms.
 It is weighed together with cheese.
 In total, the weighing scale indicates 11 kilograms.
 The same cheese is then weighed together with a milk carton.
 In total, the weighing scale indicates 3 kilograms less than before.
 What is indicated on the weighing scale now?

On moving day, Ryan is carrying his microwave oven, which weighs 5 kilograms.
 He is carrying his coffee machine at the same time.
 In total, he is carrying 11 kilograms of appliances.
 Felicia takes Ryan's coffee machine from him while carrying a blender.
 In total, she is carrying 2 kilograms less than Ryan.
 How many kilograms of appliances is Felicia carrying?

Procedure. This experiment was conducted online using the Qualtrics platform for online experiments. On the first page, the instructions read:

On the following pages, you will be presented with a series of short math problems. Your task is to solve the problems using as few operations as possible. We ask that you take enough time to read and understand each of these problems, as this is not a speed test. Remember that the goal is to solve

the problems using as few operations as possible. For every problem, we ask you to type down every operation(s) that you used to come up with the solution, even the simplest one that you can calculate mentally. For instance, the computation “ $15 - 6 - 2 = 7$ ”, should not be written as a unique operation, but broken down as “ $15 - 6 = 9$ ” and “ $9 - 2 = 7$ ”, which then count for two operations.

A different problem statement was displayed on each of the 18 following pages. We used 9 cardinal problems and 9 ordinal problems (see Table 6 and Table 7, column “Original problem statement”). Problem order was randomized between participants.

When participants had answered every problem, they were presented with a short distractor task designed to increase the rate of mistakes in the following recognition task by spacing out the solving and the recognition tasks. The distractor task consisted of three short situations in which participants had to select an explanation for a natural phenomenon among three different interpretations. For instance:

Bob, John and Lydia are trying to figure out why animals die. Here are their ideas:

- Bob: ‘Because they are mortal’
- John: ‘Because they need to leave room on earth’
- Lydia: ‘Because they die when their body stops functioning’

With whom do you agree more?

In each distractor situation, three different explanations were proposed. One displayed a circular reasoning (Bob), one was teleological (John) and the other was the closest from a scientific explanation (Lydia). The distractor task did not involve any numerical value nor did it include any theme related to one of the problem statements. When participants had chosen an explanation for the three situations, they were then presented with the unexpected recognition task. The following instructions were displayed:

In the next part of this experiment, you will be presented with a series of problem statements. Some of these problems will be strictly identical to the ones you solved in the first part of the experiment, and some will be slightly different. For each problem, a sentence will be highlighted in red. Your task will be to decide, for each problem, whether the sentence highlighted in red is the same as before or whether it has been modified. The part of the text that

is not highlighted in red will be no different in either case. Please read the problem statements entirely and take the time to understand them, as this is not a speed test.

Participants were then presented with a series of 18 problem statements to evaluate. The fifth sentence was systematically highlighted in red, and participants had to answer the question “Is the sentence highlighted in red the same as before?”. Two thirds of the problems were presented in their modified version (right column of Table 6 and Table 7); they were the focus of our analyses. In addition to these 12 target problems, 6 unmodified problems were introduced; they were identical to the ones that had been presented in the solving task (left column of Table 6 and Table 7).

Results

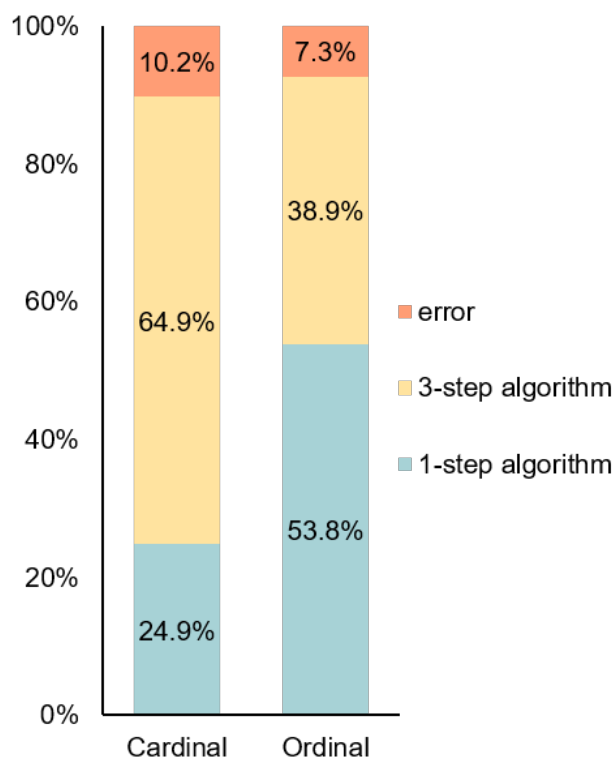


Fig. 4 Strategy distribution depending on the quantities used in the problems.

In this experiment, the algorithms used by the participants to solve the problems could always be inferred from their report of the operations they used to solve the problems, so there are no “unidentified” answers. This difference with Experiments 1 and 2 might be attributable to the monetary compensation provided to the participants on the Mechanical Turk experiment, which might have increased their incentive to

follow the instructions and write down every operation they used. As in the previous experiments, their responses were either classified as “1-step algorithm” (successful use of the shortest strategy), “3-step algorithm” (successful use of the longest strategy) or “error” (failure to use any relevant algorithm to solve the problem). The distribution of the participants’ solving strategies depending on the ordinal versus cardinal nature of the problems is described in Fig. 4. As in Experiment 1 and 2, we used a generalized linear mixed model with a binomial distribution to evaluate how likely participants were to use the 1-step algorithm on cardinal and on ordinal problems. We selected the successful use of the 1-step algorithm as the dependent variable, the semantic nature of the problems as a fixed effect and the participants as a random effect. As in the previous experiments with problems written in French, the problems written in English were considerably more likely to be solved using the 1-step algorithm when they featured ordinal quantities (53.80%) than when they featured cardinal quantities (24.85%); $z = 8.35$, $p < .001$, $R^2_{\text{GLMM}(G)} = .77$.

Regarding the recognition task, we studied how likely participants were to falsely recognize modified problems, depending on the cardinal versus ordinal nature of the problems, as well as on the solving strategies they used in the solving task (see Fig. 4). Since it was shown that participants who fail to solve a problem tend to make a random choice in a following recognition task (Hegarty et al., 1995; Thevenot, 2010), we focused our analyses on the problems that had been correctly solved with either algorithm in the solving task. We used a generalized linear mixed model with a binomial distribution to identify which factors influenced the participants’ responses on the modified problems. We used the response to the recognition task as the dependent variable, the semantic nature of the problems and the solving strategy as two fixed effects, and we accounted for each participant solving each type of problem by including a random effect for each respondent, as well as a random effect accounting for variations between problem statements. The model successfully converged with a total explanatory power of 49% (conditional R^2). Results showed that, as hypothesized, participants were more likely to incorrectly recognize the modified problems when they were ordinal problems (52.14% of false recognition) than when they were cardinal problems (16.99% of false recognition); $z = 6.54$, $p < .001$. Additionally, participants’ rate of false recognition was also dependent on which strategy they used to solve the problems in the first place: participants who successfully solved a problem were more likely to incorrectly recognize a modified

problem if they had solved the original problem with the 1-step algorithm (45.40% of false recognition) than if they had solved it with the 3-step algorithm (26.39% of false recognition); $z = 2.05$, $p < .05$. Thus, both the semantic nature of the problems and the ability to use the 1-step algorithm in the solving task influenced participants' tendency to mistake a modified sentence for an original one.

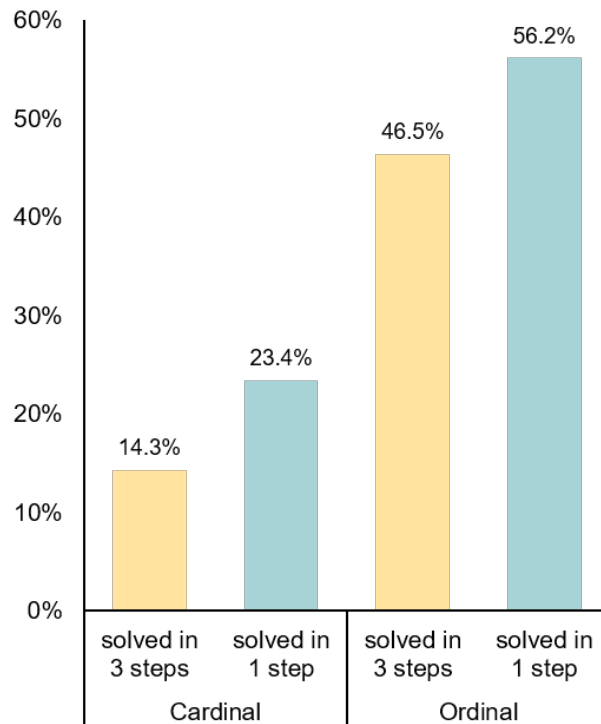


Fig. 5 Rate of erroneous recognition of altered sentences, depending on the quantities used in the problems and on the strategies used to solve them.

Discussion

This third experiment conducted with native English speakers brought substantial evidence that the distinction between cardinal and ordinal quantities has a crucial role on the representation of arithmetic word problems, displaying an influence strong enough that it shaped participants' recollection of the situations described, as well as the algorithms they used to solve the problems. Across 18 different contexts, participants' solving strategies were significantly influenced by the quantities involved in the problems, which suggests that the representations they constructed were different on cardinal and on ordinal problems. The sentence recognition task revealed that participants' memory of the different situations differed between cardinal and ordinal problems. Using a recognition task allowed us to probe participants' representation of the difference described in the problems in a new way. Results were aligned with the hypothesis that the initial encoding of the ordinal problems included

the whole-to-whole inference, whereas the initial encoding of the cardinal problems did not convey this piece of information. This suggests that participants did not perceive the isomorphism between the 18 problems, and that the representations they constructed were influenced by their non-mathematical knowledge about the quantities featured in the described situations. Irrelevant knowledge about elevators, marbles, prices or weights had a significant impact on the nature of the representation they encoded. Additionally, the ability to use the 1-step algorithm in the solving task was also a predictor of participants' rate of false recognition of the modified problems, even after the effect of the cardinal versus ordinal dimension was accounted for. In other words, using the 1-step algorithm, regardless of the nature of the quantities it involves, increases the chance to falsely recognize its modified version. An axis-based representation makes it easier to perceive that the part-to-part difference is equal to the whole-to-whole difference, and thus in return makes it harder to identify how the difference was phrased in the original problem.

General discussion

Since direct inspection of mental constructs is hardly feasible, numerous indirect routes have been proposed in the past by cognitive scientists aiming to scrutinize the representations underlying human thought processes. The various paths to study one's representation of a given idea or situation range from highly subjective methods such as verbal reports (*e.g.* Ericsson & Simon, 1980) or self-assessment questionnaires (*e.g.* Weinman, Petrie, Moss-Morris, & Horne, 1996), to more impartial measures such as reaction times (*e.g.* Rosch, 1975), operand recognition (*e.g.* Thevenot & Oakhill, 2006), relational priming (*e.g.* Bassok, Pedigo, & Oskarsson, 2008), inductive projection (Inagaki & Hatano, 1996), growing lines estimation (*e.g.* Casasanto & Boroditsky, 2008), event-related potentials (Berkum, Hagoort, & Brown, 1999), or even fMRI activations (*e.g.* Ischebeck, Schocke, & Delazer, 2009), while also including intermediate sources of information such as drawings (*e.g.* Vosniadou & Brewer, 1992), written statements (*e.g.* Pinnegar, Mangelson, Reed, & Groves, 2011), metaphors (*e.g.* Lakoff & Núñez, 2000) or gestures (*e.g.* Fuhrman & Boroditsky, 2010). In this paper, we chose to inspect the differences between the encoding of cardinal and ordinal problems by investigating the (mis)remembrance of specific problem statements.

Using unexpected recall and recognition tasks, it was shown that problems involving ordinal quantities foster a type of implicit inference that can easily be drawn from an axis-based representation, but that would be harder to make using a set-based representation. This corroborates previous findings suggesting that irrelevant, non-mathematical knowledge interferes with our encoding of numerical situations. Interestingly, previous studies have shown that ordinal representations were somewhat more adequate to solve problems, since they led to the use of a shorter solving algorithm (Gros et al., 2017) or to a higher performance rate (Gros et al., 2016, 2019). However, in this study, we showed that encoding an ordinal representation also meant being more prone to make specific mistakes in text recall and sentence recognition tasks. Participants performed better on ordinal problems in the solving tasks, but this effect was inverted in the recall and recognition tasks. This suggests that the influence of world knowledge is a double-edged sword: it is a source of inferences that can, depending on the task, prove useful or detrimental.

We have known ever since Loftus's work on the creation of false memories in long-term memory that recall and recognition can be tempered with by leading participants to represent a situation they have never lived (Loftus, 1996; Loftus & Pickrell, 1995). More recently, it has been suggested that working memory could also give rise to false memories (Abadie & Camos, 2019). Here, although we did not plant entirely false memories in the participants' minds, we led them to misremember mathematical information about the scenes described in the problems, by eliciting one of two contrasting encodings of mathematically identical situations. The fact that the use of one quantity over another was enough to predict whether participants would infer a specific relational statement between two mathematical entities of the problems suggests that our general, non-mathematical knowledge about the world has a profound impact on our mathematical reasoning. Although often overlooked, this worldly influence on mathematical reasoning is not as bizarre as it may seem. Indeed, the surface features of situations are generally correlated with their deeper principles (Bassok, Wu, & Olseth, 1995; Blessing & Ross, 1996; Gentner & Medina, 1998; Trench & Minervino, 2015). Which means that using the superficial aspects of situations to infer their deep structure is often a fruitful approach. It is only natural then that humans automatically rely on contextual clues to comprehend the situations they encounter. Despite mathematics being a field where abstraction reigns supreme (Davis, Hersh, & Marchisotto, 2011), being able to infer which mathematical notions

need be used in a specific context may be a valuable skill. For example, we tend to infer from our experience that dividing a number of apples by a number of fruit baskets may be sensible, whereas dividing a number of apples by a number of oranges is less likely to be useful in daily-life (Bassok, Chase, & Martin, 1998). However, the downside is that when the inferences we make based on our knowledge about the world conflict with the mathematical structure of a situation, strong limitations may hinder our ability to use the most efficient strategy, or to properly recall a problem statement.

In fact, the frequency with which participants inferred new relational statements only in ordinal problems provides new insights into the difficulty met by solvers in their attempts to see past the superficial dissimilarities between problems and perceive the isomorphism between cardinal and ordinal situations (Gros et al., 2015). If additional pieces of information are deduced from only one of the two possible encodings of the situation, then it makes sense that transfer from one situation to another would be especially challenging. Although systematic comparison between problem statements and reference to the deep structure of the problems might help improve transfer to some extent (Gamo et al., 2010), the question of what it takes to systematically see the deep structure of arithmetic word problems regardless of the inferences drawn from the contexts they are embedded in remains a decisive issue.

References

- Abadie, M., & Camos, V. (2019). False memory at short and long term. *Journal of Experimental Psychology: General*, 148(8), 1312-1334.
- Bassok, M. (2001). Semantic alignments in mathematical word problems. *The analogical mind: Perspectives from cognitive science*, 401-433.
- Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology*, 35, 99-134.
- Bassok, M., Pedigo, S. F., & Oskarsson, A. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 343-352.
- Bassok, M., Wu, L. L., & Olseth, K. L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory and Cognition*, 23, 354-367.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. 2014.
- 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.
- Black, J. B., & Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior*, 20(3), 267-275.
- Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3), 792.
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, 11(2), 177-220.
- Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. In W. Chase (Ed.), *Visual information processing* (pp. 383-438). New York, NY: Academic Press.
- Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, 106(2), 579-593.


- Colomé, À., & 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.
- Corbett, A. T., & Doshier, B. A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behavior*, 17(4), 479-491.
- Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, 20(4), 405-438.
- Dantzig, T. (1945). *Number the language of science. A critical survey written for the cultured non-mathematician*. New York, NY: The Macmillan Company.
- Dauben, J. W. (1990). *Georg Cantor: His mathematics and philosophy of the infinite*. Princeton University Press.
- Davis, P., Hersh, R., & Marchisotto, E. A. (2011). *The mathematical experience*. Berlin, Germany: Springer Science & Business Media.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, 87(3), 215.
- Fischer, F. E., & Beckey, R. D. (1990). Beginning kindergarteners' perception of number. *Perceptual and Motor Skills*, 70(2), 419-425.
- Frege, G. (1980). *The Foundations of Arithmetic* [1884], transl. J. Austin. Evanston, IL: Northwestern University Press.
- Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, 34(8), 1430-1451.
- Gamo, S., Sander, E., & Richard, J.-F. (2010). Transfer of strategy use by semantic recoding in arithmetic problem solving. *Learning and Instruction*, 20, 400-410.
- Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number*. Cambridge, MA, US: Harvard University Press.
- Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 65(2-3), 263-297.
- Gros, H., Sander, E., & Thibaut, J. P. (2016). "This problem has no solution": When closing one of two doors results in failure to access any. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1271-1276). Austin, TX: Cognitive Science Society.

- 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*, online first. doi: 10.3758/s13423-019-01628-3.
- Gros, H., Thibaut, J. P., & Sander, E. (2015). Robustness of semantic encoding effects in a transfer task for multiple-strategy arithmetic problems. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 818–823). Austin, TX: Cognitive Science Society.
- Gros, H., Thibaut, J. P., & Sander, E. (2017). The nature of quantities influences the representation of arithmetic problems: Evidence from drawings and solving procedures in children and adults. In R. Granger, U. Hahn, & R. Sutton (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 439–444). Austin, TX: Cognitive Science Society.
- Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, 87(1), 18.
- Inagaki, K., & Hatano, G. (1996). Young children's recognition of commonalities between animals and plants. *Child Development*, 67(6), 2823-2840.
- Ischebeck, A., Schocke, M., & Delazer, M. (2009). The processing and representation of fractions within the brain: An fMRI investigation. *NeuroImage*, 47(1), 403-413.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Harvard University Press.
- Kintsch, W., & Bates, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, 3(2), 150.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological review*, 92(1), 109.
- Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review*, 85(5), 363.
- Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and language*, 29(2), 133-159.
- Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books.

- 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.
- Loftus, E. (1996). Memory distortion and false memory creation. *Bulletin of the American Academy of Psychiatry and Law*, 24(3), 281-295.
- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25(12), 720-725.
- Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition*, 10(2), 181-187.
- Meyer, C., Barbiers, S., & Weerman, F. (2016). Order and Ordinality: The Acquisition of Cardinals and Ordinals in Dutch. In *Boston University Conference on Language Development (BUCLD) 40*. Cascadilla.
- 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.
- 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.
- Noordman, L. G., & Vonk, W. (2015). Inferences in discourse, psychology of. In *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed.) Vol. 12 (pp. 37-44). Elsevier.
- Pinnegar, S., Mangelson, J., Reed, M., & Groves, S. (2011). Exploring preservice teachers' metaphor plotlines. *Teaching and Teacher Education*, 27(3), 639-647.
- R Core Team (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192.
- Russell, B. (1919). *Introduction to mathematical philosophy*. London: 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.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans and understanding: An inquiry into human knowledge structures*. New York, NY: Hillsdale.

- Staub, F. C., & Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension* (pp. 285–305). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, 103(2), 255.
- Suppes, P. (1972). *Axiomatic set theory*. New York, NY: 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., & 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.
- Trench, M., & Minervino, R. A. (2015). The role of surface similarity in analogical retrieval: Bridging the gap between the naturalistic and the experimental traditions. *Cognitive Science*, 39(6), 1292-1319.
- Verschaffel, L. (1994). Using retelling data to study elementary school children's representations and solutions of compare problems. *Journal for Research in Mathematics Education*, 25(2), 141-165.
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24(4), 535-585.
- 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.
- 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.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358(6389), 749.

CHAPTER 4 – WHEN MASTERS OF ABSTRACTION
RUN INTO A CONCRETE WALL:
EXPERTS FAILING ARITHMETIC WORD PROBLEMS



Presentation

In Chapter 1, it was suggested that experts may be able to directly perceive a word problems' deep structure, in accordance with Chi, Feltovich, and Glaser's (1981) findings. In this chapter, we investigate this claim by testing whether experts can ignore the semantics imbued in a problem statement to abstract its mathematical structure regardless of the cover story used. We use two solvability-assessment experiments, akin to the sixth experiment of Chapter 2, to determine if experts can indeed use the 1-step algorithm on cardinal problems without first needing to engage in a semantic recoding of their initial representation.

Experiment 1 is conducted with lay adults. They are asked to evaluate, as quickly as possible, which problems can and which cannot be solved using the 1-step algorithm that is provided to them. We predict that the encoding difference between problems involving cardinal quantities and problems involving ordinal quantities is robust enough to hinder participants' ability to acknowledge the validity of the 1-step algorithm proposed. Thus, adult participants should make more mistake on cardinal than on ordinal problems, and their correct responses should require more time on cardinal problems, due to their need to engage in a semantic recoding step.

Experiment 2 attempts to replicate the findings of the first experiment with a population of experts, recruited based on their outstanding level in mathematics. We predict that general mathematical expertise is not enough to overcome the effects of semantic congruence: the initial encoding of the problems will still be influenced by the cover stories, despite participants' experience with abstract, context-independent reasoning. By demonstrating how robust the influence of world semantics on mathematical reasoning is, even among expert mathematicians, this chapter provides insights into the pervasiveness of the effects of semantic (in)congruence.

This article was published in 2019 in *Psychonomic Bulletin & Review*.

Abstract

Can our knowledge about apples, cars, or smurfs hinder our ability to solve mathematical problems involving these entities? We argue that such daily-life knowledge interferes with arithmetic word problem solving, to the extent that experts can be led to failure on problems involving trivial mathematical notions. We created problems evoking different aspects of our non-mathematical, general knowledge. They were solvable by one single subtraction involving small quantities, such as $14 - 2 = 12$. A first experiment studied how university-educated adults dealt with seemingly simple arithmetic problems evoking knowledge which was either congruent or incongruent with the problems' solving procedure. Results showed that in the latter case, the proportion of participants incorrectly deeming the problems "unsolvable" increased significantly, as did response times for correct answers. A second experiment showed that expert mathematicians were also subject to this bias. These results demonstrate that irrelevant non-mathematical knowledge interferes with the identification of basic, single-step solutions to arithmetic word problems, even among experts who have supposedly mastered abstract, context-independent reasoning.

Keywords

arithmetic reasoning · cardinality · encoding effects · ordinality · problem solving

Introduction

Is $14 - 2 = 12$ always obvious? Most third graders know the basics of addition and subtraction (Carpenter & Moser, 1984), and solving elementary arithmetic operations is no big deal from this point onwards. We learn from an early age that operations such as $14 - 2 = 12$ are always valid, no matter whether one is subtracting apples, cars, or smurfs. However, our claim is that adults whose mathematical knowledge is unquestionable, even outstanding, sometimes fail to solve arithmetic problems admitting a single-step solution such as $14 - 2 = 12$, when their knowledge about the entities subtracted interferes with the mathematical structure of the problem.

This prediction arises from a growing body of literature suggesting that the daily-life, non-mathematical world knowledge one has about the objects an arithmetic word problem refers to might influence their mathematical representation of the problem and their subsequent choice of a solving strategy. For example, Bassok, Wu, and Olseth (1995) showed that being trained to solve a permutation problem was not always helpful to solve analogous problems. The authors demonstrated that slight, mathematically irrelevant, changes in the semantic relations linking the objects mentioned in the cover stories (e.g., computers assigned to secretaries versus secretaries assigned to computers) led to significant performance differences. Subsequent research has shown that non-mathematical semantic information related to the entities described in a problem influences lay solvers' performance (Bassok, Chase, & Martin, 1998; Gros, Sander, & Thibaut, 2016; Thevenot & Barrouillet, 2015; Verschaffel, De Corte, & Vierstraete, 1999; Vicente, Orrantia, & Verschaffel, 2007) as well as strategy choice (Gamo, Sander, & Richard, 2010; Gros, Thibaut, & Sander, 2017) and transfer (Gros, Thibaut, & Sander, 2015) on arithmetic word problems. Most of the available evidence regarding this issue has been collected with children and non-expert adults, on problems that were not straightforward (e.g., complex permutation problems). Building on this literature, we propose to go further and show that irrelevant aspects of what we call *world semantics* (the non-mathematical knowledge about the world that is evoked by the entities described in a specific problem statement) can also mislead experts in mathematics on problems involving basic arithmetic notions, despite them being considered experts in abstract, context-independent reasoning (Dehaene, 2011). We call this proposal the “world semantics view”.

Despite stemming from the aforementioned literature, this claim that world semantics could exert such a pervasive influence and threaten even the highest levels of mathematical expertise is rather innovative, as it challenges the commonly held view in the expertise literature regarding experts' proficiencies. This expertise view notably considers that experts identify what has been described as the "deep structure" of the problem (Chi, Feltovich, & Glaser, 1981), its "principle" (Ross, 1987), its "objective mathematical structure" (Bassok, 2001) or its "problem space" (Newell & Simon, 1972). This deep structure is independent of the semantics imbued in the problem statement, and as such it is the foundation of experts' abstract, context-independent reasoning about the problem. Indeed, since by definition mathematics is not empirical and manipulates abstract symbols rather than real-life objects (Davis, Hersh, & Marchisotto, 2011; Russell, 1903), mathematical experts should ignore irrelevant information associated with the entities on which numbers and algorithms operate. They should perceive the deep structure of arithmetic problems that can be solved by simple subtractions (i.e., involving small quantities such as $14-2$), no matter whether they calculate the price of an apple, the height of a smurf or the speed of a car. Furthermore, experts are known to show exceptional performance in domain-related tasks (Chi, 2006), they stand out in their ability to generate problem solutions (De Groot, 1965), to detect relevant problem features (Lesgold et al., 1988), to monitor their own comprehension (Chi, 1978) and to qualitatively analyze the task at hand (Voss, Greene, Post, & Penner, 1983) (see Chi, 2006 for a review of experts' proficiencies). These former studies do not predict that the semantics conveyed by the problem statement could interfere with the experts' understanding of the problems' mathematical structure.

We performed two experiments to show that, contrarily to this expertise view - but in accordance with the world semantics view - arithmetic problems admitting a single-step solution might pose a challenge to mathematical experts. We presented participants with a series of isomorphic problems involving two numerical values. Crucially, for each problem, a solution was provided (a single subtraction between the problem's two numerical values), and participants' task was to evaluate its validity. By varying the semantic, non-mathematical information evoked by the problem statements (e.g., use of an elevator versus a weighing scale, reference to marbles being won versus years passing by, mention of hamburger prices versus statues' heights, etc.), we intended to show that even math experts are exposed to a

deleterious influence of the non-mathematical knowledge evoked by the problem statement.

Our world semantics view predicts that university students (Study 1) – and math experts (Study 2) – will more often fail to recognize the proposed solution when it conflicts with the non-mathematical knowledge about the world evoked by the entities featured in the problem statement than when the solution is consistent with it. Furthermore, it predicts that a recoding process, akin to re-representation (Davidson & Sternberg, 2003; Vicente et al., 2007) is necessary when a problem's initial encoding leads to a dead end. Indeed, when the semantic content of a problem statement leads participants to interpret the situation in a way that is not compatible with the problem's solution, then it becomes necessary to build a new representation of the situation congruent with the solution. When successfully performed, such a recoding process should result in longer response times for correct answers conflicting with the problems' world semantics.

Study 1

Methods

Participants. We recruited 85 adults (50 women, Mean age = 23.35, $SD = 7.82$) in the Paris region. All had attended university (Mean length of university curriculum = 2.85 years, $SD = 1.18$), but none majored in mathematics. Considering the low complexity of the math problems involved, participants' curriculum was a clear indicator that they possessed the mathematical expertise required to solve the problems. Sample size was determined using uncertainty and publication bias correction on results from a previous study (Gros et al., 2016), following Anderson, Kelley, and Maxwell's recommendations (2017).

Materials. Our materials were inspired by Gamo et al. (2010) who showed that problems with the same formal mathematical structure are nevertheless preferentially solved with one of two available solving strategies, depending on the semantic content of the problem. Consider the weight problem in Table 1: this problem can be solved through two strategies. One is a 3-step algorithm consisting in calculating the weight of each individual dictionary to compute the weight of the stack of dictionaries Lola is carrying: $14 - 5 = 9$; $5 - 2 = 3$; $9 + 3 = 12$. The other one is a 1-step algorithm that requires understanding that since Lola and Joel carry the same Spanish dictionary,

calculating the weight of each book is unnecessary. Since the German dictionary is 2 kilograms lighter than the Russian dictionary, the weight difference between Joel's and Lola's books is of 2 kilograms as well: $14 - 2 = 12$.

Table 1 Two isomorphic problems sharing the same mathematical structure but evoking different aspects of our knowledge about the world.

Weight problem	Duration problem
Joe takes a Russian dictionary weighing 5 kilograms.	Tom took painting classes for 5 years.
He also takes a Spanish dictionary.	He stopped taking the classes at the age of 14.
In total, he is carrying 14 kilograms of books.	Lucy started taking painting classes at the same age as Tom.
Lola takes Joe's Spanish dictionary and a German dictionary.	She took classes for 2 years less than him.
The German dictionary weighs 2 kilograms less than the Russian dictionary.	How old was Lucy when she stopped taking painting classes?
How many kilograms of books is Lola carrying?	

The duration problem in Table 1 has the same mathematical structure and can be solved using the same solving procedures. However, Gamo et al. (2010) showed that the two solving procedures are not randomly distributed across the two types of problems. Participants favor the 3-step algorithm on problems like the dictionary one (called *cardinal* problems) and the 1-step algorithm on the second type of problems (called *ordinal* problems). This strategy use imbalance was our starting point. Gamo et al. (2010) and Gros et al. (2017) showed that the differences in the world semantics evoked by the problems resulted in different spontaneous encodings of the situations, from which this imbalance originated¹ (see Fig. 1 for a description of this effect). Since cardinal and ordinal problems shared the same structure featuring the same parts and wholes presented in the same order with the same numerical values, the imbalance in strategy use could only be attributed to the variations of the semantic content of the problem statements. Additionally, when considering the correct answers on either

¹ Although the explanation of this effect is not the purpose of the present paper, the authors suggest that because our world knowledge about dictionaries says we can stack them with no specific order, they evoke a representation of the total as a combination of subsets, which they call a cardinal representation. A similar reasoning can be held for weights or prices defined as object properties (Gros et al., 2017). On the other hand, using the 1-step algorithm requires participants to build a re-representation of the problem that is not based on a “combination of subsets”, which makes computing the weight of the Spanish dictionary unnecessary. By contrast, some problems seem to emphasize the ordinal nature of the values featured and afford a representation of the numerical values on a continuous axis. For example, we spontaneously encode durations on a timeline, which makes it easier for school children and lay adults to notice that the numerical difference between the two distinct parts is equal to the difference between the two totals (Gamo et al., 2010). A similar reasoning can be held for height or floor problems (Gros et al., 2017). Thus, using the 1-step algorithm is more straightforward for ordinal than for cardinal problems (see Fig. 1).

algorithm there was no significant difference in adults' performance between cardinal and ordinal problems, which indicates that the strategy imbalance was not a matter of problem difficulty (Gros et al., 2017).

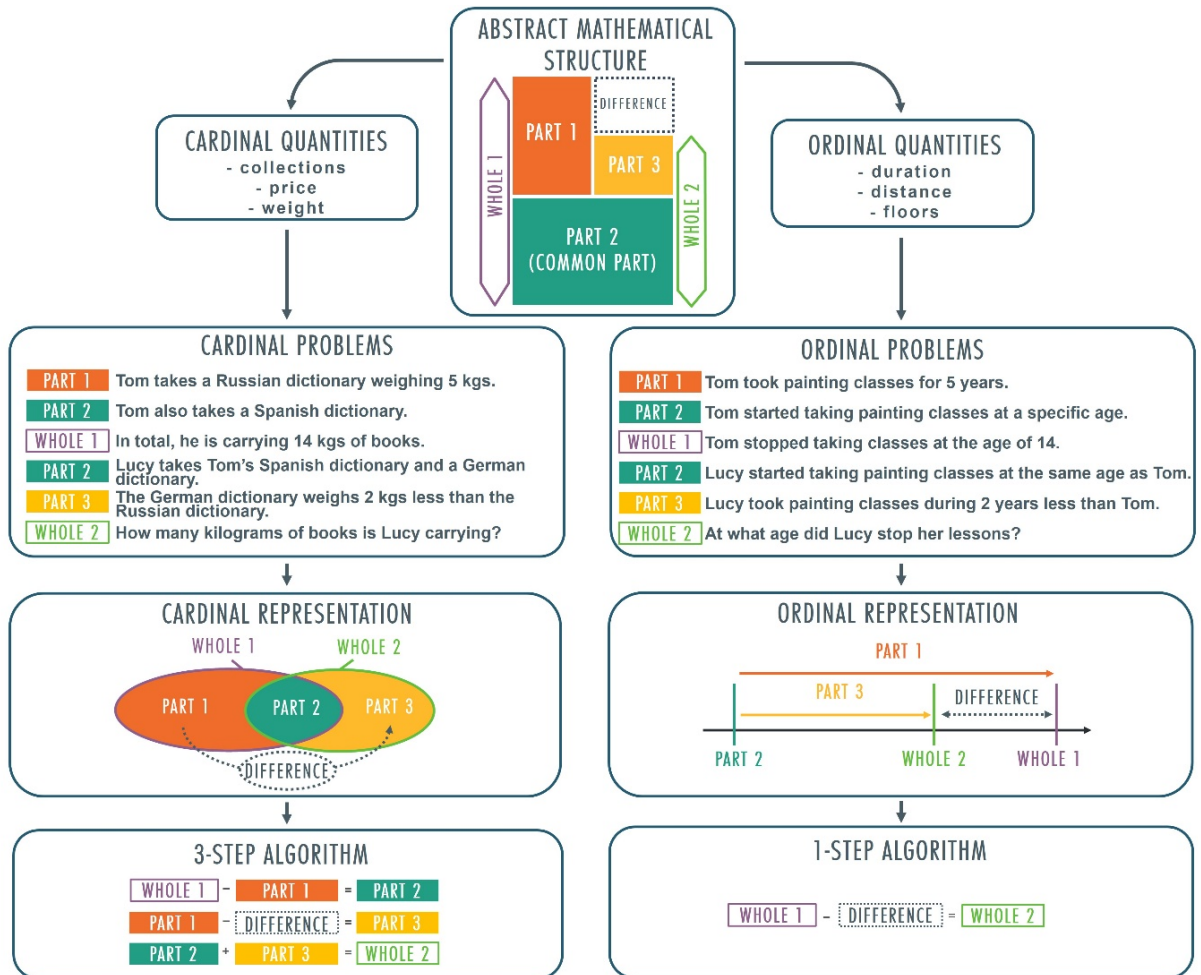


Fig. 1 Implementation of the mathematical structure with ordinal versus cardinal quantities, leading to different problem statements, representations, and strategy use.

Gros et al., (2017) have shown that most adults encode collection, price, and weight problems as cardinal representations whereas they encode duration, distance and floor problems as ordinal representations. We modified their problems and removed the value of *Part 1* so that the 3-step strategy could not be used (see Table 2). Consequently, the only solution left was the 1-step strategy, which required using the values of *Whole 1* and of the *Difference* (see Fig. 1). The materials are available online (https://osf.io/fxgqh/?view_only=ed1374ef4d204c90a0cb03a30cb0a099).

Ordinal problems were 333.5 characters long on average (SD = 38.37) and cardinal problems were 304 characters long on average (SD = 44.94). This length difference was not statistically significant ($t(10) = 1.18, p = .26$, paired t-test). Crucially,

for each problem, participants were presented with the correct 1-step solution (e.g. “ $14 - 2 = 12$; Jolene has 12 marbles”). Participants’ task was to decide whether the provided solution worked, or whether there was no solution to the problem. Due to the already established imbalance in strategy use between problems evoking a cardinal encoding and problems evoking an ordinal encoding (Gamo et al., 2010; Gros et al., 2017), we assumed that the measure of participants’ ability to use the only remaining strategy on problems evoking different aspects of world semantics would be an effective assessment of the robustness of these effects.

Table 2 Example of target problems used in the study. Changes introduced from Gros et al.’s problem statements are italicized in the table for the sake of clarity, but they were not made apparent in the experiment. Translated from French.

Cardinal target problems	Ordinal target problems
<p>Paul has <i>a certain amount of</i> red marbles. He also has blue marbles. In total, Paul has 14 marbles. Jolene has as many blue marbles as Paul, and some green marbles. She has 2 green marbles less than Paul has red marbles. How many marbles does Jolene have?</p> <p>In the store, Anthony wants to buy a ruler <i>costing a certain price</i>. He also wants a notebook. In total, that will cost him 14 dollars. Julie wants to buy the same notebook as Anthony, and an eraser. The eraser costs 2 dollars less than the ruler. How much will Julie have to pay?</p> <p>Joe takes a Russian dictionary weighing <i>a certain weight</i>. He also takes a Spanish dictionary. In total, he is carrying 14 kilograms of books. Lola takes Joe's Spanish dictionary and a German dictionary. The German dictionary weighs 2 kilograms less than the Russian dictionary. How many kilograms of books is Lola carrying?</p>	<p>Sofia travelled <i>for a certain time</i>. Her trip started during the day. Sofia arrived at 14 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?</p> <p>Slouchy Smurf is <i>a certain height</i>. He climbs on a smurf table. He now attains the height of 14 centimeters. Grouchy Smurf climbs on the same table as Slouchy Smurf. Grouchy Smurf is 2 centimeters shorter than Slouchy Smurf. What height does Grouchy Smurf attain when he climbs on the table?</p> <p>Katherine took the elevator and went up <i>a certain number of floors</i>. She left from the floor where the gym is. She arrived to the 14th floor. Yohan also took the elevator from the floor where the gym is. He went up 2 floors less than Katherine. What floor did Yohan arrive to?</p>

The world semantics hypothesis predicts lower performances on cardinal than on ordinal problems, even among experts, because cardinal problems would require a re-representation of the situation when the only solution available is the 1-step

algorithm. By contrast, ordinal problems should be easier to solve because participants' spontaneous encoding facilitates the use of the 1-step algorithm. Since university-educated adults can be considered experts in solving subtractions such as $14 - 2 = 12$, and since the deep structure of a problem is identical regardless of the objects involved, this prediction could not be made without the world semantics view, especially when participants only need to check the validity of the proposed solution. Additionally, we predict that recoding a situation initially encoded as a combination of subsets (such as a cardinal encoding) into a representation in terms of states and transitions between states (such as an ordinal encoding) is a costly process, requiring a longer response time. Although our hypotheses only regard solvable problems, we also included unsolvable distractor in the materials, so that the correct answer would not always be "This problem can be solved". Among those distractors the value of *Whole 1* was removed instead of the value of *Part 1*, which rendered the problems unsolvable with either algorithm.

Procedure. Participants answered the questions using three keyboard keys on a 17" laptop. Instructions stated that "Some of the problems can be solved using the values provided, while other problems cannot be solved with the available information. Your task is to tell apart problems that can be solved from problems that cannot. Answer as quickly as you can, although being correct is more important than being fast."

Participants were presented with 6 target problems that were only solvable with the 1-step algorithm: 3 cardinal and 3 ordinal problems. An equal number of distractors was introduced to fulfill subjects' expectations regarding the uniform distribution of yes/no answers. Problem order, cover stories and numerical values were randomized between participants. The value of *Whole 1* was comprised between 11 and 15, *Whole 2* between 5 and 9, and the *Difference* was either 2 or 3.

We used a segmented self-presentation procedure displaying the text line by line on the screen when participants pressed the spacebar. Below, a question appeared: "Given the data provided, is it possible to find the solution?" followed by two possible choices: "A) No, there is not enough information to find the solution.", "B) Yes, and the following solution is correct:" (followed by, in the case of the marble problem: " $14 - 2 = 12$. Lucy has 12 marbles in total"). A solution was proposed for each problem, and it was up to the participants to assess whether it was valid or whether the problem was unsolvable.

Results

Data collected for both studies are available online (https://osf.io/fxgqh/?view_only=ed1374ef4d204c90a0cb03a30cb0a099). The dependent variable was the proportion of correct answers for solvable problems (see Fig. 2). Because multiple binary data points were recorded in a repeated design (each participant provided a binary answer to 3 ordinal and 3 cardinal solvable problems), the use of repeated measures ANOVA was deemed inappropriate and replaced by a mixed model (Hector, 2015). We used a generalized linear mixed model with a binary distribution, with the cardinal versus ordinal semantic nature of the problems as a fixed factor, and participants as a random effect. In line with our hypothesis, lay adults performed significantly better on ordinal (81.18%) than on cardinal problems (46.67%); $z = 7.84$, $p < .001$, $R^2_{\text{GLMM}(G)} = .29^2$. Additionally, looking at individuals' response patterns showed us that 65.9% of the participants made fewer mistakes on ordinal than on cardinal problems, 11.8% made no mistakes at all, 15.3% made the same number of mistakes in cardinal and in ordinal problems and only 7.1% made more mistakes on ordinal than on cardinal problems.

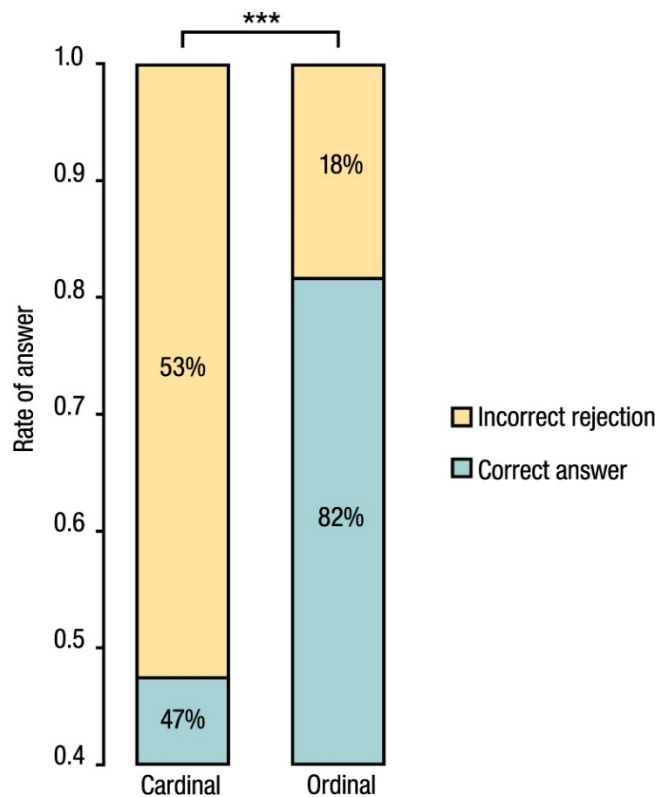


Fig. 2 Adults' answers distribution. *** $p < .001$

² Conditional R^2 are reported in lieu of η^2 for the mixed models in this paper, since no satisfying method is currently available to estimate effect sizes on mixed models (Westfall, Kenny, & Judd, 2014).

Further analyses were conducted on participants' response times (RTs) on solvable problems that had been successfully identified as such by the participants (see Fig. 3). Because the number of correct answers could vary from 0 to 6 for each participant, the number of RT data points varied accordingly, and the use of repeated measures ANOVA was again deemed inappropriate (Hector, 2015). A linear mixed model with subjects as a random effect and semantic nature of the problems as a fixed factor showed that participants took more time to correctly solve cardinal ($M = 34.05$, $SD = 18.78$) than ordinal problems ($M = 26.85$, $SD = 12.49$), $\chi^2(1) = 29.14$, $p < .001$, $R^2_{LMM(C)} = .44$. Additionally, we studied the participants' individual response patterns to identify whether different participant profiles existed. For each participant, we computed the difference between their mean RTs on correctly solved cardinal and ordinal problems (see Fig. 4) and we performed Hartigan's dip test for unimodality versus multimodality on the resulting distribution (Hartigan & Hartigan, 1985). The analysis failed to reject the null hypothesis that participants' responses came from a unimodal distribution ($D = .028$, $p = .94$), thus providing no empirical ground to assume that the distribution of response times was multimodal.

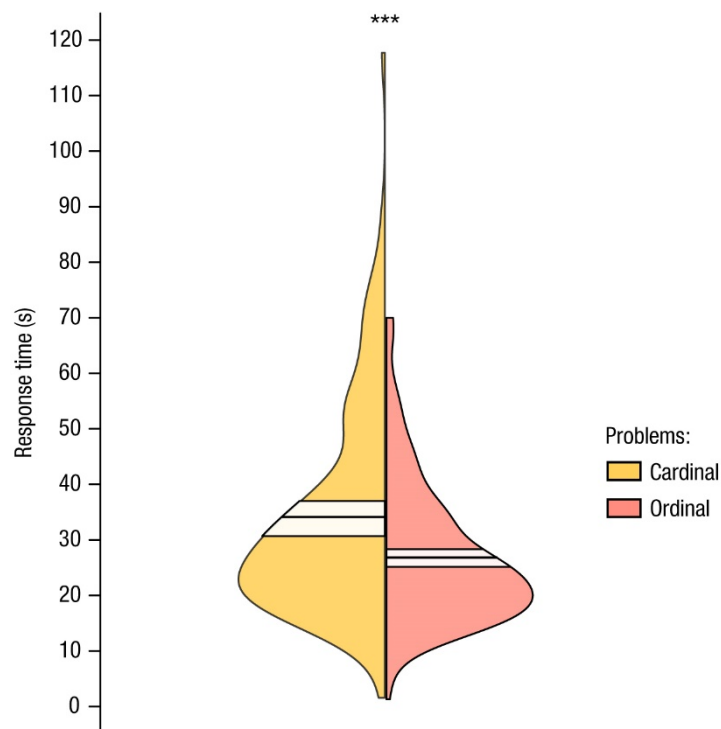


Fig. 3 Violin plot of adults' RTs on correctly identified solvable problems. Middle bars indicate mean RTs; upper and lower bars indicate margins of .95 confidence intervals. *** $p < .001$

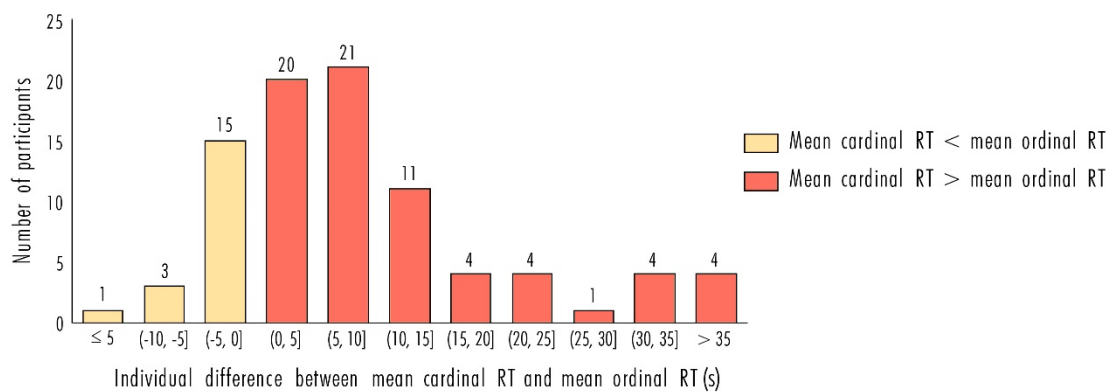


Fig. 4 Distribution of individual differences between cardinal RT and ordinal RT on correctly solved problems. Bins below the zero value indicate participants whose ordinal RT were higher than their cardinal RT on average, whereas bins above zero indicate participants whose ordinal RT were lower than their cardinal RT on average.

Discussion

The difference in performance between cardinal and ordinal problems indicates that despite their expertise regarding basic subtractions, the adults' answers were significantly influenced by the semantic content of the problem statements. This confirms previous results obtained with the “complete” version of the problems that could be solved either with the 3-step algorithm or with the 1-step algorithm (Gamo et al., 2010; Gros et al., 2017). Here, we showed that the strategy imbalance observed in these previous studies was not an effect of mere preference for one strategy over another, but an actual impossibility to identify the relevance of the 1-step algorithm on cardinal problems, as attested by the fact that on these problems, more than half of the participants rejected a perfectly valid solution, despite only needing to check its validity. Regarding RTs, the fact that correct answers took more time on cardinal problems suggests that recognizing the solution to a problem evoking aspects of world semantics seemingly incompatible with the solution required an extra processing step. This is also supported by the fact that there was no significant difference in length between cardinal and ordinal problems. This is in line with the recoding process we predicted. These results show that the semantic content of a problem can prevent university-educated adults from recognizing a simple subtraction as the solution to a problem whose mathematical structure is undoubtedly within their level of expertise. We designed a second study to identify whether such effects would remain with expert mathematicians, known to be especially accustomed to abstract reasoning.

Study 2

Methods

Participants. We recruited 25 experts (2 women, Mean age = 23.59, $SD = 2.81$) who had successfully passed the entrance exam of the Science section at the École Normale Supérieure (ENS Ulm) in Paris. This exam is considered as the most demanding one in France, with an entrance rate of 2.02% among university-educated participants (“SCEI Statistics”, 2017). The ENS ranked second in Times Higher Education’s World University Rankings 2016-2017 for Best Small University (Bhardwa, 2017). Although the population sample was smaller than in the first study due to the number of graduates from École Normale Supérieure being limited, sample size was deemed sufficient using uncertainty and publication bias correction on results from a previous study (Gros et al., 2016), following Anderson et al.’s recommendations (2017).

Materials and Procedure. Materials and procedure were identical to that of Study 1.

Results

As in Study 1, we analyzed the proportion of correct answers on solvable problems (see Fig. 5) with a generalized linear mixed model. Experts had a higher success rate on ordinal (94.67%) than on cardinal problems (76.00%); $z = 2.99$, $p = .0028$, $R^2_{\text{GLMM}(c)} = .25$. Additionally, a comparison with Study 1 showed that Study 2 experts’ performance (85.33%) was significantly higher than Study 1 adults’ performance (63.92%), which was another confirmation of their outstanding expertise in mathematics; $z = 4.49$, $p < .001$, $R^2_{\text{GLMM}(c)} = .33$. Looking at individuals’ response patterns also indicated that 52.0% of the participants made fewer mistakes on ordinal than on cardinal problems, 36.0% made no mistakes at all, 4.0% (1 participant) made the same number of mistakes in cardinal and in ordinal problems and only 8.0% made more mistakes on ordinal than on cardinal problems.

Analyses were conducted on participants’ RTs for correctly identified solvable problems (see Fig. 6). As in Study 1, we used a linear mixed model that showed that experts took significantly more time to correctly solve cardinal problems ($M = 26.58$, $SD = 14.03$) than ordinal problems ($M = 19.45$, $SD = 8.18$), as predicted by our world semantics hypothesis; $\chi^2(1) = 18.65$, $p < .001$, $R^2_{\text{LMM}(c)} = .37$. Unsurprisingly, experts’ RTs on correct answers were significantly shorter ($M = 22.63$, $SD = 11.68$) than in

Study 1 ($M = 29.50$, $SD = 15.48$); $\chi^2(1) = 7.68$, $p = .0056$, $R^2_{\text{LMM}(C)} = .46$. As in Study 1, the computation of individual differences in RTs between cardinal and ordinal problems showed no sign of multimodality (see Fig. 7), and Hartigan's dip test for unimodality versus multimodality failed to reject the null hypothesis of unimodality ($D = .048$, $p = .96$).

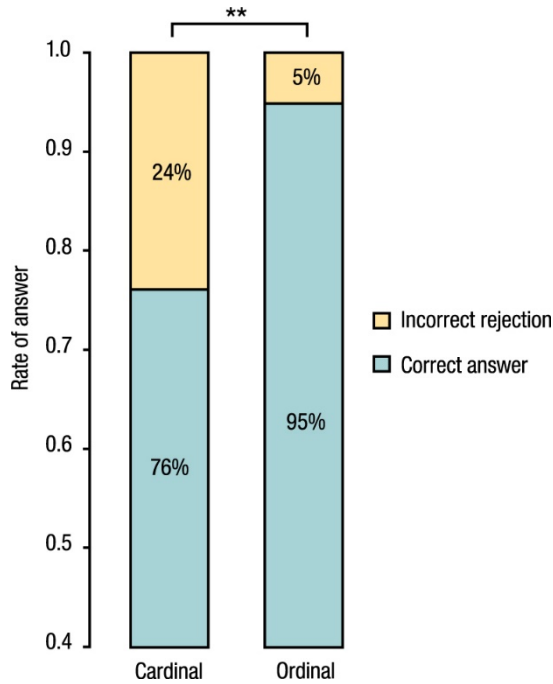


Fig. 5 Experts' answers distribution. ** $p < .01$

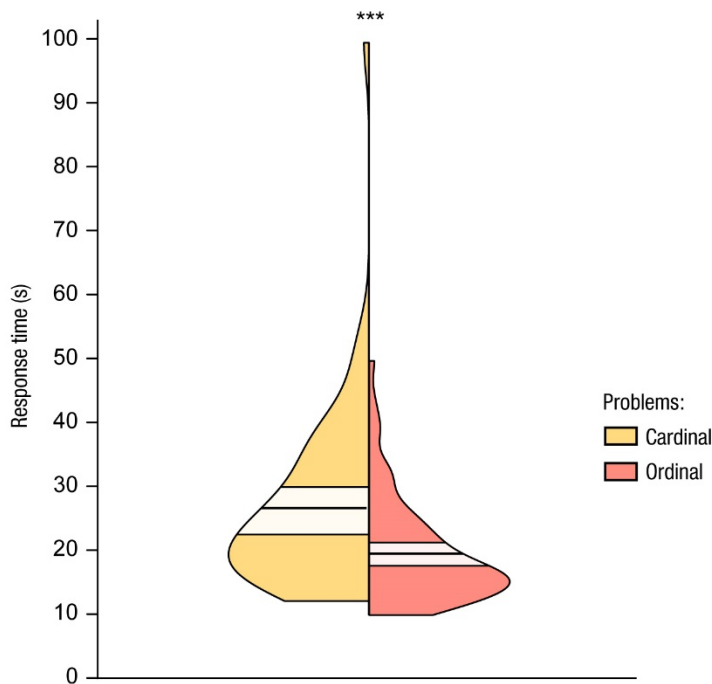


Fig. 6 Violin plot of experts' RTs on correctly identified solvable problems. Middle bars indicate mean RTs; upper and lower bars indicate margins of .95 confidence intervals. *** $p < .001$

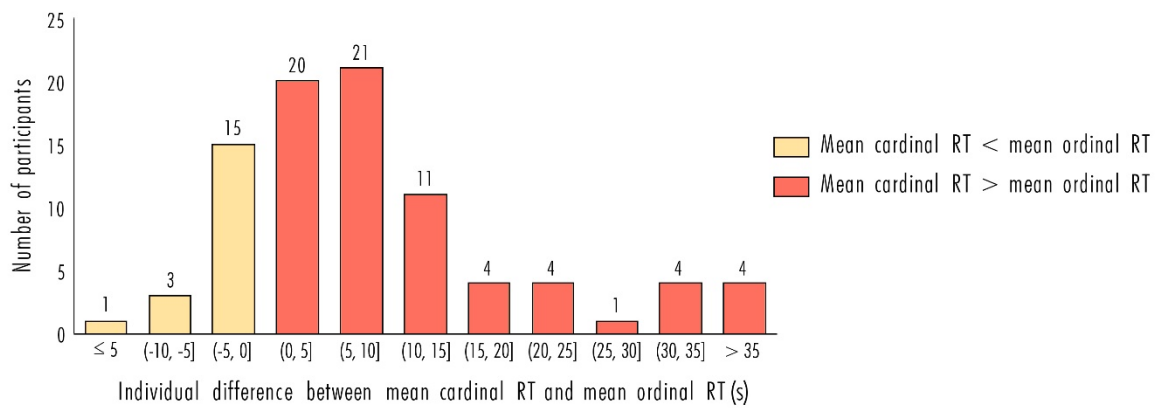


Fig. 7 Distribution of individual differences between cardinal RT and ordinal RT on correctly solved problems. Bins below the zero value indicate participants whose ordinal RT were higher than their cardinal RT on average, whereas bins above zero indicate participants whose ordinal RT were lower than their cardinal RT on average.

Discussion

Despite their superior performances, high-level mathematicians were still significantly influenced by world semantics. Their performance dropped significantly on cardinal problems, and correct answers required more time on average on cardinal than on ordinal problems. Therefore, despite their proficiency in abstract mathematical reasoning, expert mathematicians failed to disregard irrelevant non-mathematical information when solving the problems, as hypothesized.

General discussion

In this paper, we sought to demonstrate that irrelevant aspects of our non-mathematical knowledge evoked by the semantic content of a problem statement can lead both adults and mathematics experts to encode the problem in such a way that they would erroneously consider valid solutions as incorrect. Indeed, participants failed to identify the solvability of subtraction problems admitting a single-step solution significantly more often when the world semantics they evoked conflicted with the relevant mathematical information, than when the two were congruent. Additionally, correct answers took more time in the conflicting than in the congruent case for both populations, suggesting that the initial spontaneous representation triggered by the semantic content of the problem statement had to be recoded. Although they achieved higher performances overall, high-level experts still rejected several perfectly valid solutions: they fell prey to robust effects of world semantics that current theories of expertise do not account for.

There is a much larger body of literature describing in what terms experts excel in their field than there are studies revealing experts' shortcomings. However, as Chi (2006, p. 23) stressed, "it is equally important to understand how experts fail", which was one of the goals of this paper. A few limitations have already been shown to occasionally affect experts' excellence (see Chi, 2006 for a review). For instance, experts' proficiencies are limited to their domain of expertise (Ericsson & Lehmann, 1996) and they lack adaptability to irregular situations whose structures differ from what they expect (Sternberg & Frensch, 1992). They have even been shown to gloss over details (Voss, Vesonder, & Spilich, 1980), which paradoxically suggests that they should be good at ignoring surface properties unrelated to the formal structure of the problems. More recent works have even hinted at biases slowing down experts within their own domain of expertise (Goldberg & Thompson-Schill, 2009; Obersteiner, Van Dooren, Van Hoof, & Verschaffel, 2013). However, we believe none of these accounts would have predicted our results, since they do not explain how mathematically irrelevant contextual information may significantly hinder experts' abstract reasoning on problems within their very field of expertise, to the extent that they would not identify the validity of the solution handed out to them. Here, mathematical experts failed to do what they are good at: engaging in abstract reasoning on concrete entities to find a single-step solution. Our results suggest that when mathematical knowledge and world semantics conflict with one another, masters of abstraction can run into a concrete wall.

This effect is understandable since world semantics and mathematical knowledge often (although not always) naturally align with each other, which explains how some superficial cues are highly correlated with deeper principles (Bassok, Pedigo, & Oskarsson, 2008; Blessing & Ross, 1996). It follows that solvers rely on those cues at all levels and tend to make mistakes when world and mathematical semantics do not align. Overall, it seems that these effects of semantic (in)congruence between world semantics and mathematical knowledge have been greatly undermined on the account of mathematics being an inherently abstract domain in which rules and concepts are valid independently from the objects they are applied to. Our results show how prevalent the influence of world knowledge is on arithmetic reasoning, even among the individuals who should be the least subject to it. This suggests that experts will never be completely freed from the influence of

world knowledge; having an outstanding level in mathematics is not enough to systematically perceive that $14 - 2 = 12$.

Acknowledgments

We sincerely thank Pernille Hemmer and two anonymous reviewers whose insightful feedback helped improve and clarify this manuscript. We also acknowledge gratefully Pierre Barrouillet, Katarina Gvozdic and Maxime Maheu for helpful comments on previous versions of this work.

References


- 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.
- Bassok, M. (2001). Semantic alignments in mathematical word problems. In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 401–433). Cambridge, MA: MIT Press.
- Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology, 35*(2), 99–134.
- Bassok, M., Pedigo, S. F., & Oskarsson, A. T. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(2), 343–352.
- Bassok, M., Wu, L. L., & Olseth, K. L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory and Cognition, 23*, 354–367.
- Bhardwa, S. (2017). International Student Table 2017: Top 200 Universities. *Times Higher Education*.
- Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(3), 792–810.
- Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15*, 179–202.
- Chi, M. T. H. (1978). Knowledge structures and memory development. In R.S. Siegler (Ed.) *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 121–130). Cambridge: Cambridge University Press.
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.

- Davidson, J. E., & Sternberg, R. J. (Eds.). (2003). *The psychology of problem solving*. New York, NY: Cambridge University Press.
- Davis, P., Hersh, R., & Marchisotto, E. A. (2011). *The mathematical experience, Study edition*. Boston, MA: Birkhäuser.
- De Groot, A. D. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.
- Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47(1), 273–305.
- 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.
- Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental “roots” in mature biological knowledge. *Psychological Science*, 20(4), 480–487.
- Gros, H., Sander, E., & Thibaut, J. P. (2016). “This problem has no solution”: When closing one of two doors results in failure to access any. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1271–1276). Austin, TX: Cognitive Science Society.
- Gros, H., Thibaut, J. P., & Sander, E. (2015). Robustness of semantic encoding effects in a transfer task for multiple-strategy arithmetic problems. In D. C., Noelle, R., Dale, A. S., Warlaumont, J., Yoshimi, T., Matlock, C. D., Jennings, & P. P., Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 818–823). Austin, TX: Cognitive Science Society.
- Gros, H., Thibaut, J. P., & Sander, E. (2017). The nature of quantities influences the representation of arithmetic problems: Evidence from drawings and solving procedures in children and adults. In R. Granger, U. Hahn, & R. Sutton (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp 439–444). Austin, TX: Cognitive Science Society.
- Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 13(1), 70–84.
- Hector, A. (2015). *The new statistics with R: An introduction for biologists*. Oxford, UK: Oxford University Press.

- Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser, M. J. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum
- Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-Hall.
- Obersteiner, A., Van Dooren, W., Van Hoof, J., & Verschaffel, L. (2013). The natural number bias and magnitude representation in fraction comparison by expert mathematicians. *Learning and Instruction, 28*, 64–72.
- Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(4), 629–639.
- Russell, B. (1903). *Principles of mathematics*. Cambridge, UK: Cambridge University Press.
- "SCEI Statistics", (2017). Retrieved from <http://www.scei-concours.fr/statistiques.php>
- Sternberg, R. J., & Frensch, P. A. (1992). On being an expert: A cost-benefit analysis. In *The psychology of expertise* (pp. 191–203). New York, NY: Springer.
- Thevenot, C., & Barrouillet, P. (2015). Arithmetic word problem solving and mental representations. In R. Cohen Kadosh, & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 158–179). Oxford, UK: Oxford University Press.
- 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.
- Vicente, S., Orrantia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology, 77*(4), 829–848.
- Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. In *Psychology of learning and motivation* (Vol. 17, pp. 165–213). New York, NY: Academic Press.
- Voss, J. F., Vesonder, G. T., & Spilich, G. J. (1980). Text generation and recall by high-knowledge and low-knowledge individuals. *Journal of Verbal Learning and Verbal Behavior, 19*(6), 651–667.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*, 2020–2045.

CHAPTER 5 – ARE CONTENT EFFECTS OUT OF
TEACHERS' SIGHT? AN EYE-TRACKING STUDY
OF ARITHMETIC WORD PROBLEM SOLVING



Presentation

Chapter 5 has a twofold ambition. First, while Chapter 4 asked the question of how well experts in the field might perform when faced with semantic incongruence, this new chapter focuses on a different kind of expertise: that of teachers. Indeed, it can be argued that mathematical expertise in general does not necessarily mean expertise to solve arithmetic word problems. In fact, word problems are a peculiar exercise, common in schools, but rarely encountered in their scholastic form during adulthood. Expert mathematicians have proficient Content Knowledge regarding mathematics, but their Pedagogical Content Knowledge (Shulman, 1986) regarding arithmetic word problems may be less extensive than that of teachers. Did our experts have too little experience with arithmetic word problems to use the full range of their mathematical expertise? To answer this question, we conducted a new study with secondary school teachers, of mathematics and other subjects, and preservice teachers.

Experiment 1 attempts to replicate the findings of Chapter 4 among a population of secondary school teachers. A total of 32 math teachers and 90 teachers with no experience teaching math are presented with a solvability-assessment task in which they are instructed to take as long as they need to decide which problems can be solved with the proposed algorithm. We predict that even math teachers will fall prey to the effects of semantic (in)congruence, despite their expertise regarding word problems.

Second, we used eye-tracking data to better characterize the encoding distinction accounting for the performance difference between cardinal and ordinal problems. By registering participants' eye movements when solving problems similar to the ones used in Experiment 1, we were able to get a finer understanding of the nature of the differences between cardinal and ordinal problems, as well as of what happens when participants manage to discard an initially incongruent representation. Experiment 2 was conducted with 50 pre-service teachers, asked to evaluate the solvability of 18 arithmetic word problems. We predict that the difference between cardinal and ordinal problems will be observed on success rates, response times, visit durations, backward eye movements, and pupil dilation.

Abstract

Recent evidence suggests that general, non-mathematical knowledge about the entities described in an arithmetic word problem may interfere with its encoding, to the extent that even math experts may struggle to solve problems embedded in a semantically incongruent context. In this study, we used behavioral and eye tracking measures to investigate how the use of specific quantities (weights, prices, collections) may foster a cardinal representation of the numbers mentioned in a problem, whereas other quantities (durations, heights, number of floors) may favor an ordinal representation instead. In Experiment 1, we assessed the pervasiveness of the cardinal-ordinal distinction in a population of secondary school teachers ($N = 122$) in a solution validity assessment task. We found that even math teachers accustomed to the didactic specificities of mathematical word problems made more errors and took a longer time to solve semantically incongruent problems. In Experiment 2, we recorded the eye movements of pre-service teachers ($N = 50$) engaged in a similar task. We compared participants' gaze patterns on cardinal and ordinal isomorphic problems to gather insights into their encoded representations of the problems. On problems featuring cardinal quantities, we found that specific sentences describing elements relevant in a cardinal understanding of the problems but irrelevant otherwise were looked at longer, and were the focus of a higher number of backward eye movements. Additionally, increase in pupil dilation on correctly solved cardinal problems supported the idea that participants need to engage in a semantic recoding process when in a situation of semantic incongruence. Overall, the results support the growing line of evidence that daily-life knowledge about the quantities featured in a problem may interfere with its solving by promoting one of two competing representations.

Keywords

mathematical cognition · eye tracking · arithmetic word problems · problem solving · representation · encoding · mathematics · mental models

Introduction

Semantic determinants of arithmetic word problem solving

Mathematical word problems are infamously difficult, and many a student have struggled with the delicate exercise consisting in applying abstract mathematical notions to concrete, daily-life examples (Cummins, Kintsch, Reusser, & Weimer, 1988; Daroczy, Wolska, Meurers, & Nuerk, 2015; Fayol, Barrouillet, & Camos, 1997; Lewis & Mayer, 1987; Nesher & Teubal, 1975; Riley, Greeno, & Heller, 1983; Stern, 1993; Verschaffel, Greer, & De Corte, 2000). By bringing reality into the classrooms, word problems are meant to prepare the students to face the mathematical situations they will inevitably encounter in their everyday lives (Dewolf, Van Dooren, Hermens, & Verschaffel, 2015; Pollak, 1969). But what makes some mathematical word problems so hard to solve? A significant part of their complexity can be attributed to the need to extract relevant mathematical information from a non-mathematical text, which implies to deal simultaneously with the linguistic and mathematical complexity of the word problems (Thevenot, 2017). Neither mathematical fluency on its own nor text comprehension skills are enough to find the solution to a mathematical word problem, and several lines of work have looked at the interaction between linguistic and numerical factors in the interpretation of mathematical word problems (Daroczy et al., 2015; Thevenot & Barrouillet, 2015; Verschaffel et al., 2000).

Notably, the issue of the underlying representations accounting for the strategies developed by students to solve the problems they encounter has been a recurring question in the literature. It has for example been proposed that students use problem schemata: abstract general frames, stored in long-term memory, that are implemented with a given problem's numerical values to find its solution (Kintsch & Greeno, 1985; Riley et al., 1983; Schank & Abelson, 1977). Another competing approach has suggested that the construction of mental models (Johnson-Laird, 1983) of the situations depicted in the problems constituted an intermediate step for the solvers before reducing it to its mathematical gist and using the resulting solving algorithm (Staub & Reusser, 1995). More recently, it has been suggested that general semantic knowledge about the entities featured in a problem could interfere with its solving process, by the means of an interpreted structure describing one's interpretation of the situation depicted in the problem (Bassok, 2001). In this paper, we intend to evaluate the predictions of a complementary approach suggesting that

an initial semantic representation is encoded based on the problem statement and on the solver's prior knowledge about the quantities it features. An important aspect of this approach is that it notably predicts that a fruitless encoding of a given problem statement may sometimes be semantically recoded in an attempt to overcome a dead end and find the solution to an arduous problem (Gamo, Sander, & Richard, 2010; Gros, Sander, & Thibaut, 2019). We propose to take a look at the role of prior knowledge on the encoding, recoding and solving of arithmetic word problems by studying the perception of cardinality and ordinality among teachers and pre-service teachers, using behavioral and eye tracking data.

Cardinal versus ordinal encoding: investigating the constructed representations

In 1978, Gelman and Gallistel argued that learning how to count implied to master a series of “counting principles” governing the very activity of counting. Among those, learning the “stable-order-principle” means that a child needs to understand that the list of words used to count has to be used in a fixed and repeatable order: numbers follow each other in a definite succession that does not fluctuate. By putting such emphasis on the notion of order, this principle introduces the ordinal meaning of numbers. On the other hand, learning about the “cardinal principle” entails that a child makes the connection between the number name allocated to the final element that is being counted, and the total number of elements in the collection. In other words, when counting entities one by one, the final number name being used refers to the total number of entities in the set. Those two principles define the difference between the ordinal and the cardinal meanings of numbers, which are central to the notion of number itself (Fuson, 1988). Following this work, several studies have investigated the development of counting among children to identify the age at which they master the ordinal principle (Fischer & Beckey, 1990; Miller, Major, Shu, & Zhang, 2000; Miller, Marcovitch, Boseovski, & Lewkowicz, 2015) as well as the cardinal principle (Bermejo, 1996; Condry & Spelke, 2008; Le Corre & Carey, 2007; Sarnecka & Lee, 2009; Wynn, 1992). More recently, works have been conducted to compare the developmental trajectories of these two sides of counting, thus stressing the importance of the difference between cardinality and ordinality in numerical cognition (Colomé & Noël, 2012; Meyer, Barbiers, & Weerman, 2016; Wasner, Moeller, Fischer, & Nuerk, 2015). However, these studies have focused on the first steps of learning how to count, and the importance that the distinction between cardinality and

ordinality still holds once counting is mastered has received scant attention in the field.

A recent line of work has aimed to fill this gap, by targeting the cardinal and ordinal representations of arithmetic word problems among older children and adults. Preliminary work on this question was undertaken by Gamo et al. (2010), who found that children tended to approach isomorphic problems differently depending on the semantic nature of the quantities they used. More precisely, they showed that students' choice of solving algorithms varied between number-of-element problems, price problems and age problems. They suggested that this discrepancy could be attributed to differences in the semantic encoding of the problems: while number-of-element problems and price problems feature unordered elements that tend to be represented as sets and subsets, age problems are more easily represented along an axis (a timeline) and the apparent order between the age values facilitates the use of a different solving strategy. This distinction between ordered and unordered quantities was framed in terms of ordinal and cardinal encodings, and the idea that quantities emphasizing the cardinal aspect of numbers led to different representations that quantities underlining their ordinal aspect was introduced.

In an attempt to investigate this distinction in a systematic way, new arithmetic word problems were created using different types of quantities. Gros, Thibaut and Sander (2017) proposed problems designed specifically to evoke one of two encodings: they used collection (number of elements), price, and weight problems to evoke a cardinal representation, and they used duration, height and number-of-floors problems to evoke an ordinal representation. Fig. 1 provides a graphical summary of the hypothesis they tested. The problems all shared the same abstract mathematical structure (Fig. 1, box 1.), but they were implemented either with cardinal quantities (Fig. 1, box 2.a) or with ordinal quantities (Fig. 1, box 3.a.). Consider for example the cardinal problem reproduced in Fig. 1, box 2.b.: by mentioning weights assigned to discrete entities such as dictionaries, this problem is expected to emphasize the cardinal aspect of numbers, and thus to elicit a cardinal encoding of the situation (Fig. 1, box 2.c.). This representation fosters the idea that to find the weight of the stack of books that Lola is carrying (*Whole 2*), one needs to add up the weight of the Spanish dictionary (*Part 2*) and the weight of the German dictionary (*Part 3*). This representation is thus semantically congruent with a 3-step algorithm (Fig. 1, box 3.c.) consisting in calculating the value of *Part 2* ($Whole\ 1 - Part\ 1 = Part\ 2$), and adding it

to the value of *Part 3* ($Part\ 1 - Difference = Part\ 3$), to find the solution to the problem ($Part\ 2 + Part\ 3 = Whole\ 2$).

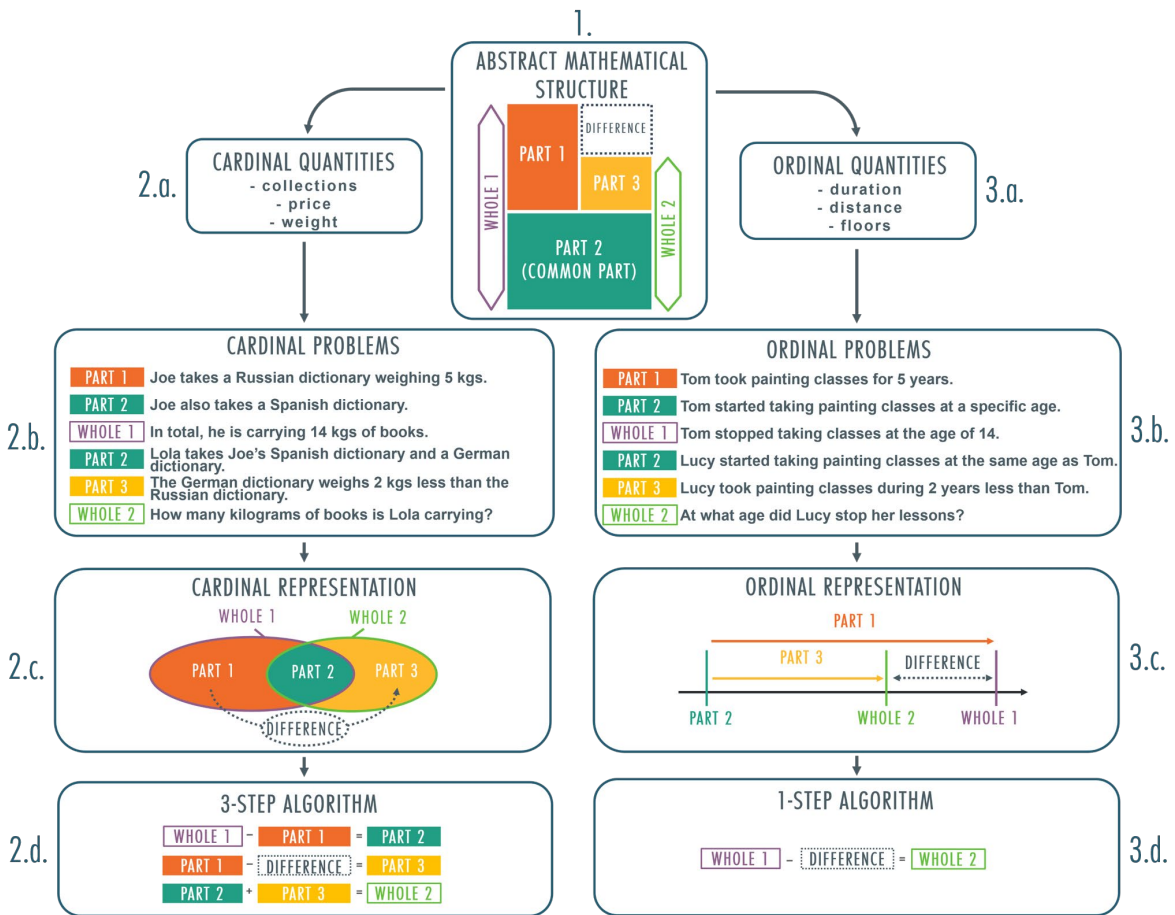


Fig. 1 Implementation of the mathematical structure with ordinal versus cardinal quantities, leading to different problem statements, representations, and strategy use. Adapted from “When masters of abstraction run into a concrete wall: Experts failing arithmetic word problems,” by H. Gros, E. Sander, & J.-P. Thibaut, 2019, *Psychonomic Bulletin & Review*, online first.

On the other hand, the duration problem reproduced in box 3.b. describes a situation that can easily be represented along an axis (such as a timeline), and it is thus thought to evoke an ordinal encoding (Fig. 1, box 3.c.). This representation facilitates the understanding that since Tom and Lucy started taking painting classes at the same age, and since Lucy took classes for 2 years less than Tom, then she stopped attending the classes when she was 2 years younger than him. Thus, this inference that the difference between *Part 1* and *Part 3* is equal to the difference between *Whole 1* and *Whole 2* makes it easier to use a 1-step algorithm to find the value of *Whole 2*: $Whole\ 1 - Difference = Whole\ 2$ (see Fig. 1, box 3.d.). The same pattern was hypothesized to happen for every problem used in the study. Depending on the cardinal versus ordinal nature of the quantities used in the problems, participants were thought to construct a different encoding of the situation, which led them to one of the two

possible solving algorithms. Drawing production elicited by Gros et al. (2017) supported the claim that participants constructed different representations depending on whether the quantities in the problems emphasized the cardinal nature of numbers (higher number of set-based drawings) or whether they emphasized the ordinal nature of numbers (higher number of axis-based drawings). Additionally, participants' report of the algorithms they used to solve the problems showed that they tended to use the 1-step algorithm more often on problems with ordinal quantities than they did on problems with cardinal quantities, despite being explicitly asked to solve the problems using as few operations as possible.

To evaluate the robustness of these encoding effects, Gros et al. (2019) proposed a modified version of these problems, in which the value of *Part 1* was not provided, to prevent participants from using the 3-step algorithm to solve the problems. For example, the sentence "Joe takes a Russian dictionary weighing 5 kgs" was replaced by "Joe takes a Russian dictionary weighing a certain weight", and the sentence "Tom took painting classes for 5 years" was replaced by "Tom took painting classes for a certain number of years". It thus became impossible to use the 3-step algorithm (Fig. 1, box 2.d.) since that required knowing the value of *Part 1*, and the only algorithm left to solve the problems was the 1-step algorithm (Fig. 1, box 3.d.). Note that this transformation of the problem statement turned it into a single-step subtraction problem, the solution of which consisted in subtracting the lower value from the higher one. Gros et al. created a solution-assessment task, in which these problems were presented accompanied by their solution, and participants had to decide whether the solution was correct or whether the problems could not be solved.

The idea was that participants would have no trouble solving the ordinal problems, since their ordinal encoding naturally led them to use the 1-step algorithm, but cardinal problems should be more troublesome. Indeed, a cardinal encoding fostering the calculation of *Part 2* and *Part 3* to find the value of *Whole 2* would result in a dead end, and it was thus predicted that participants would need to engage in a semantic recoding of the situation to understand the relevance of the 1-step algorithm. It was thus predicted that cardinal problems would lead to a higher number of failures, and that participants who nevertheless managed to solve the cardinal problems would require a higher amount of time to do so, due to the semantic recoding step necessary to construct a new, more appropriate representation of the situation. They presented this task to lay adults and to expert mathematicians and found that in both cases their

expertise was not enough to prevent the influence of the cardinal versus ordinal distinction: participants made more errors and took longer to solve cardinal problems. In the current study, we intend to build on this experimental paradigm to evaluate the influence of such encoding effects on a population more mindful of the didactic stakes behind scholastic exercises such as arithmetic word problem solving, and to get a finer understanding of the interpretative mechanisms at play using an eye tracking setup.

Eye tracking as an index of reasoning processes

It is established that the study of eye movement is informative with regards to what is being attended to in a given situation (e.g. Buswell, 1935; Just & Carpenter, 1980; Rayner, 1998; Yarbus, 1967). In text comprehension, an extensive line of work has used eye tracking to study the shifts of attention from one word to another, and the underlying thought processes (e.g. Clifton, Staub, & Rayner, 2007; Hyönä, Lorch, & Rinck, 2003; Just & Carpenter, 1987; Kaakinen & Hyönä, 2008; Rayner, Chace, Slattery, & Ashby, 2006; Rayner & McConkie, 1976; Van der Schoot, Vasbinder, Horsley, Reijntjes, & van Lieshout, 2009). Overall, it is generally admitted that eye fixations are correlated to the cognitive processing of information (Andra et al., 2015; Latour, 1962; Orquin & Mueller Loose, 2013). In fact, from an educational standpoint, there has been an increasing amount of literature using eye tracking to better understand students' learning processes (Lai et al., 2013; Scheiter & van Gog, 2009; van Gog & Scheiter, 2010). As stated by Andra et al. (2015, p. 241), "the merit from a didactic perspective is that we can examine how and which information students are attending to".

In the study of mathematical reasoning, eye tracking has also been used to pinpoint the integration of relevant information while performing calculations, looking at mathematical representations or solving math problems (Beitlich, Obersteiner, & Reiss, 2015; Bolden, Barmby, Raine, & Gardner, 2015; Curtis, Huebner, & LeFevre, 2016; Green, Lemaire, & Dufau, 2007; Knoblich, Ohlsson, & Raney, 2001; Merkley & Ansari 2010; Obersteiner & Tumpek, 2016; Schneider et al., 2008; Winoto, Tang, Huang, & Chen, 2017; Zhu, Luo, You, & Wang, 2018). However, despite the abundant literature using eye movements to investigate mathematical reasoning and the extensive research analyzing text comprehension using eye tracking, a surprisingly low number of studies have resorted to this methodology to understand mathematical *word* problem solving (Strohmaier, Tatsidou, & Reiss, 2018).

In fact, ever since De Corte and Verschaffel's (1986) seminal work on the matter more than thirty years ago, we are aware of less than a dozen studies who looked at mathematical word problem solving using eye movement recording. For example, De Corte, Verschaffel and Pauwels (1990) used eye movement recording to discriminate between the initial read-through of arithmetic word problems and the subsequent time spent rereading parts of the problem statement. Then, Verschaffel, De Corte, and Pauwels (1992) evaluated the validity of Lewis and Mayer's (1987) Simulation Model's predictions by recording students' total fixation times on the first two relational sentences in the statement of arithmetic word problems. They notably showed that students' longer response times on problems featuring relational terms inconsistent with their solving algorithms were due to a longer time spent on the initial reading of the first two problem sentences.

Similarly, Hegarty, Mayer and Green (1992) compared high-accuracy and low-accuracy students' reading patterns when solving arithmetic word problems. Their fixation analysis revealed that low-accuracy student's longer response times were due to numerous rereads of the problems after their initial read-through. In the same line, Hegarty, Mayer and Monk (1995) identified differences between successful and unsuccessful problem solvers' strategies in terms of eye fixations on numbers and relational terms in the problem statements, as well as in terms of total number of regressive eye movements. Later, Van der Schoot, Bakker Arkema, Horsley and van Lieshout (2009) focused on regressive eye movements to evaluate how the strategies of successful and less successful problem solvers differed when solving different types of arithmetic word problems. Finally, Dewolf et al. (2015) used fixation analysis to investigate how often students looked at representational illustrations accompanying standard and non-standard mathematical word problems.

Of note, Strohmaier, Lehner, Beitlich and Reiss (in press) recently used global measures of eye movements to investigate complex, unstructured mathematical problems including context information, text, pictures and other sources of information. They notably showed that the more difficult items were read with shorter fixations, more saccades and more regressions than the easier problems. Using the same type of stimuli in a cross-linguistic study, Strohmaier, Schiepe-Tiska, Chang, Müller, Lin and Reiss (in press) notably showed that Taiwanese students solved the problems using shorter fixations and a higher proportion of regressions than German students.

However, to the best of our knowledge, most of the research conducted on more classical arithmetic word problems using eye tracking methodology has focused on fixation times, with some works counting backward eye-movements to identify specific solving strategies. In our study, we intend to use both of these metrics to get a finer understanding of the differences between cardinal and ordinal problems, as well as a third one selected to evaluate participants' effort in the task: pupil dilation.

Pupillometry concerns the measure of pupil dilation over time. Its use in research was initiated by Hess and Polt (1964), who found that pupil tended to dilate when individuals were asked to solve multiplication non word problems of increasing difficulty. Subsequent works discovered that pupil diameter increased with memory load (Granholm, Asarnow, Sarkin, & Dykes, 1996; Kahneman & Beatty, 1966; Peavler, 1974; see Goldinger & Papesh, 2012, for a review) and with task demand in general (Beatty, 1982; Janisse, 1977; Kahneman, 1973). Further works confirmed that task-evoked pupillary response was a robust – although indirect – measure of cognitive load (Beatty & Lucero-Wagoner, 2000; Just, Carpenter, & Miyake, 2003), which makes it a valuable index to evaluate participants' effort variations when solving arithmetic word problems.

Current study

In this study, we set out to gain a better understanding of the fundamental difference between situations emphasizing the cardinal aspect of number, and situations underlining their ordinal aspect instead. Building on previous works showing that the distinction between cardinal and ordinal problems could elicit different solving strategies among children, lay adults and expert mathematicians, we strove to build a finer characterization of the cognitive mechanisms involved in the process. First, we decided to get a teacher's perspective on these problems: where mathematicians failed, maybe math teachers can overcome? It can be argued that the expert mathematicians recruited in Gros et al. (2019) have a specific expertise in abstract mathematical reasoning that might have fallen short of the peculiar exercise imposed to them in the form of arithmetic word problems. Despite being intended to prepare students to use mathematical notions in real life (Verschaffel et al., 2000), the odd nature of arithmetic word problems may be enough to throw off mathematicians who haven't opened a school textbook in years. Thus, since a secondary math teacher's expertise surely encompasses the didactic stakes behind the use of word problems,

can we expect a different outcome? Or, to put it differently: where Content Knowledge failed, can Pedagogical Content Knowledge (Shulman, 1986) prevail?

The first experiment was designed to answer this question, and to ask another one: if teaching math in secondary school is not enough to perceive the isomorphism between cardinal and ordinal problems, then what makes these problems so different that they confound most individuals who try their hand at solving them? The second experiment was designed to tackle this question. Using eye-tracking methodology, we asked pre-service teachers to solve the very problems that puzzled our experts. We collected 5 distinct metrics of their reasoning processes in the hope of reaching a finer-grained understanding of the differences between the encoding of cardinal and ordinal problems. Response times, performance rates, fixation times, regression counts and pupil dilation were analyzed in light of what we believe the cardinal versus ordinal distinction entails in terms of representational processes. Following Gamo et al.'s (2010) findings according to which one can be trained to semantically recode a fruitless representation, we took a special interest in pinpointing the conditions for such a recoding to happen. Ultimately, we believe that understanding such recoding processes is a crucial step to foster transfer of learnt strategies between semantically dissimilar contexts.

Experiment 1

Experiment 1 was designed to evaluate the robustness of the differential interpretative effects reported between cardinal and ordinal problems among secondary teachers. We modelled the experiment on the one challenging expert mathematicians in Gros et al. (2019), with one main difference: contrary to the instructions received by the math experts, the teachers were not told to solve the problems as fast as possible. In fact, they were explicitly told to take as long as they needed to solve the problems, in order to give them every opportunity to avoid mistakes and correctly identify the solvable problems. In accordance with previous studies showing the stability of the effect of semantic congruence among diverse populations, we made the hypothesis that the results from Gros et al. (2019) would be replicated among the secondary teachers, regardless of whether they were math teachers or teachers in other fields. In other words, we expected their success rate to drop on cardinal problems compared to ordinal problems, and we expected correctly solved cardinal problems to take up a longer time than correctly solved ordinal problems, due to the necessity

for the teachers to engage in a semantic recoding of their initial, inappropriate representation.

Methods

Participants. The participants were 122 secondary teachers (78 women and 44 men, $M = 35.85$ years, $SD = 7.67$) attending a State mandatory teacher training program at the University of Geneva. Among them, 90 had never taught mathematics at any level, while 32 had taught mathematics in secondary school for at least one year. Participants volunteered after giving informed consent, in exchange for course credit. All the participants spoke French fluently and none had previously participated in any similar experiment.

Materials. The materials for this experiment were the same as those created in Gros et al. (2019): 12 arithmetic word problems sharing the same abstract mathematical structure. Each participant was presented with a random selection of 6 target solvable problems: 3 of those featured cardinal quantities (1 collection problem, 1 price problem and 1 weight problem) and 3 featured ordinal quantities (1 duration problem, 1 height problem and 1 floors problem). We used a within-subject design to allow for within-subject comparisons between performance on cardinal and on ordinal problems. In addition, we introduced 6 unsolvable filler problems that were similar to the target problems but could not be solved with any algorithm, due to the value of *Part 1* being present and the value of *Whole 2* being missing, contrary to the target problems. Order of target and filler problems was randomized between participants. All the problems were written in French. Cardinal and ordinal problems did not significantly differ in number of words; $t(16) = 1.37$, $p = .19$, independent t-test. The problems were isomorphs and the numerical values used were randomized across problems.

Procedure. This experiment was conducted online using the Qualtrics platform for online experiments. After agreeing to and signing the voluntary participation form on the first screen, the instructions were displayed. They stated that “You will be presented with a series of arithmetic problems. Some of the problems can be solved using the values provided, while other problems cannot be solved with the available information. Your task is to tell apart problems that can be solved from problems that cannot. This is not a speed test: take your time to carefully read and understand each of these problems”. Problem order, cover stories and numerical values were randomized between participants. The value of *Whole 1* was comprised between 11

and 15, *Whole 2* between 5 and 9, and the Difference was either 2 or 3. On the next screen, the first problem was displayed. The presentation of the problems was self-paced: a new screen appeared each time a participant answered and clicked “next”.

Results

Success Rate. We recorded participants’ answers to the problems to evaluate whether the results from Gros et al. (2019) could be replicated among teachers. As in this study, we analyzed participants’ rate of correct answers and response times (RTs) on the cardinal and ordinal target problems. The dependent variable was the binary outcome indicating failure or success on solvable problems. Since each participant gave a binary answer to 6 cardinal problems and 6 ordinal problems, we used a generalized linear mixed model with a binomial distribution to account for the repeated measures in the experimental design. We used the cardinal versus ordinal nature of the problems as a fixed factor, the experience teaching math as another fixed factor, and participants as a random effect. The overall model successfully converged and had a total explanatory power of 17.16% (conditional R²). In line with Gros et al.’s findings on expert solvers, there was a main effect of the cardinal versus ordinal nature of the problems on participants’ success rate. In fact, teachers performed significantly worse on cardinal (58.74%) than on ordinal problems (86.26%); $z = 7.83$, $p < .001$. Similarly, there was a main effect of participants’ experience teaching math ($z = 4.82$, $p < .001$), meaning that participants who had already taught math performed higher on average (88.54% success) than participants who never did (66.73% success). There was no interaction between these two fixed factors ($z = 0.59$, $p < .56$).

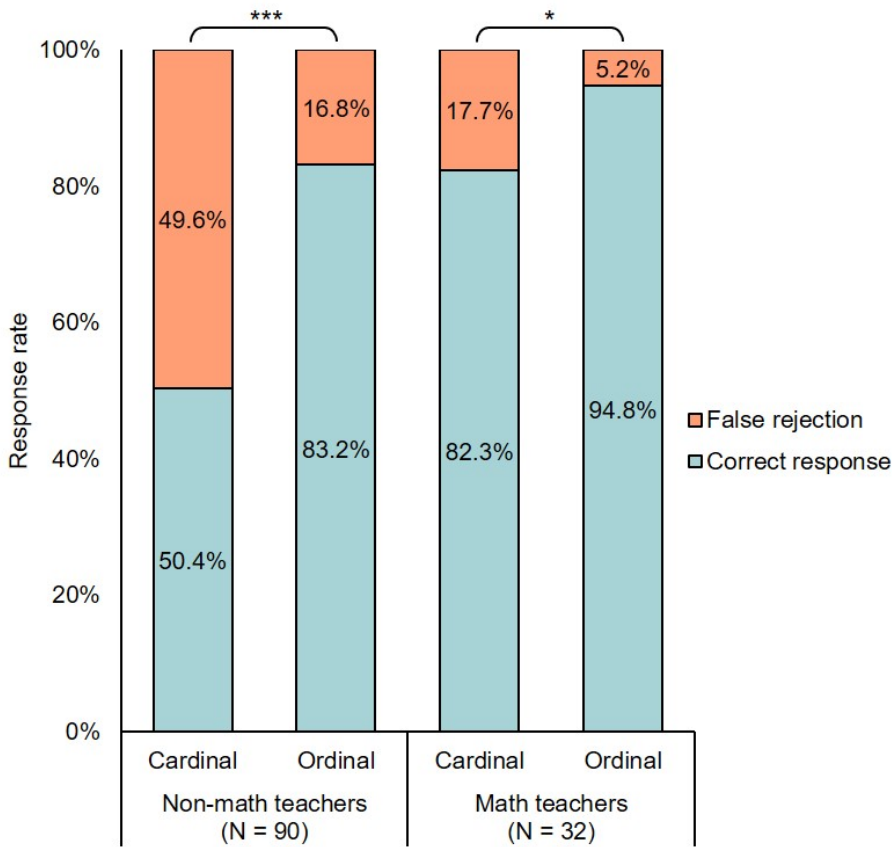


Fig. 2 Teachers' answer distribution depending on the cardinal versus ordinal nature of the problems. *** $p < .001$, * $p < .05$.

Since our hypothesis was that both groups would be influenced by the semantic difference between cardinal and ordinal problems, we conducted an additional analysis focusing on the performance difference within each group (see Fig. 2). As expected, teachers who had never taught math performed significantly better on ordinal target problems (83.21%) than on cardinal target problems (50.37%); $z = 7.83$, $p < .001$. Similarly, and despite their higher performances overall, math teachers also fell prey to the distinction between cardinal and ordinal problems: their performances on cardinal target problems (82.29%) were significantly lower than on ordinal target problems (94.79%); $z = 2.57$, $p < .05$. The results from Gros et al. (2019) regarding success rates on cardinal and ordinal problems were thus replicated in this experiment, among both groups of participants.

Response Times. Following the semantic recoding hypothesis developed in Gros et al. (2019), we investigated the RTs of correctly solved problems (see Fig. 3). We expected that correct responses on cardinal problems would require a higher amount of time than on ordinal problems, due to the need of an extra semantic recoding step to find the solution. We used Tukey's method to identify and remove 37 outliers

ranged above and below 1.5 interquartile range. We analyzed participants' response times using a linear mixed model with the cardinal versus ordinal nature of the problems as a fixed factor, the participants' experience teaching math as another fixed factor and the participants themselves as a random effect. The overall model successfully converged and had a total explanatory power of 36.86% (conditional R^2). Within this model, an ANOVA using Satterthwaite's method for estimation of degrees of freedom revealed that there was a significant effect of the cardinal versus ordinal nature of the problems on the response times on correctly solved problems ($F(1) = 19.56, p < .001$). Which indicates that correctly solving a cardinal problem required more time on average ($M = 42.88, SD = 23.79$) than correctly solving an ordinal problem ($M = 35.65, SD = 19.72$), in accordance with Gros et al.'s (2019) results. There was no main effect of the experience teaching math, meaning that math teachers' response times ($M = 42.47 \text{ s}, SD = 22.04$) did not differ significantly from that of teachers of other subjects ($M = 36.59 \text{ s}, SD = 21.24$); $F(1) = 3.78, p = .054$.

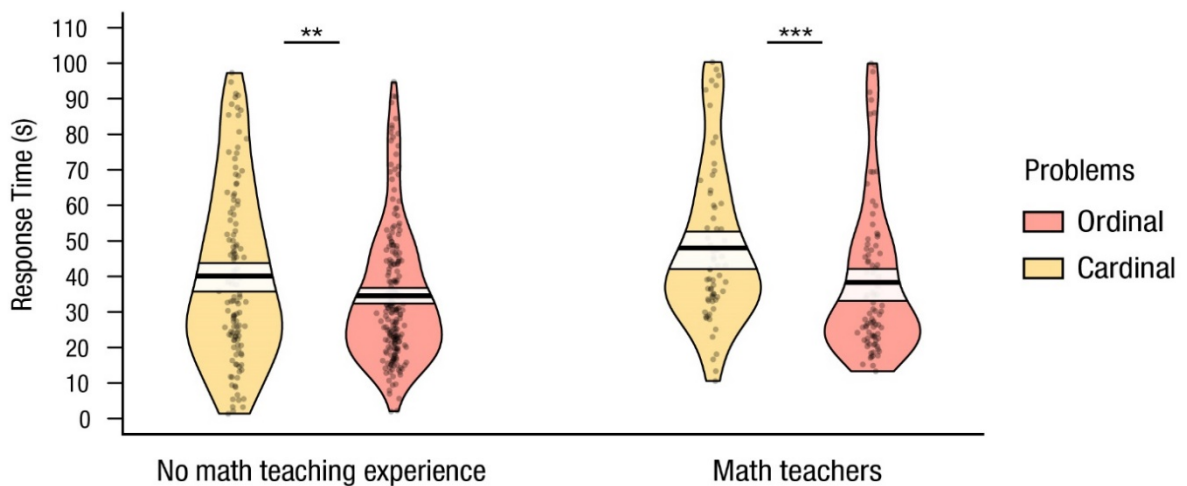


Fig. 3 Pirate plot of response times on cardinal and ordinal problems, depending on math teaching experience. Middle lines indicate mean RT, upper and lower lines indicate 95% confidence interval margins. *** $p < .001$, ** $p < .01$.

Since our hypothesis was that both groups would be influenced by the semantic difference between cardinal and ordinal problems, we conducted a follow-up analysis of RTs, focusing on the difference within each group. As expected, despite their higher success rate, math teachers also needed to perform an extra recoding step, as indicated by their longer RTs on cardinal problems ($M = 48.00 \text{ s}, SD = 21.97$) than on ordinal problems ($M = 38.32 \text{ s}, SD = 21.29$); $F(1) = 12.79, p < .001$. As expected, teachers who had never taught math also needed an extra step to solve the cardinal problems, as indicated by their faster response on ordinal target problems ($M = 34.55$

s, SD = 18.99) than on cardinal target problems ($M = 40.13$ s, $SD = 24.36$); $F(1) = 7.81$, $p < .01$.

Discussion

In this first experiment, we replicated Gros et al.'s (2019) findings among two populations: secondary teachers who never taught math, and secondary teachers with at least one year of experience teaching math in secondary schools. As expected, participants in both groups had a harder time identifying the solution to the cardinal problems than to the ordinal problems, despite the isomorphic structure of the problems. Almost half of the time, non-math teachers were not able to identify the conformity of the simple subtraction that was proposed to them to solve cardinal problems: they dismissed the problems as unsolvable, despite being provided with its solution. Despite their obvious expertise in mathematical reasoning and arithmetic word problem solving, math teachers were also influenced by the cardinal versus ordinal semantics imbued in the problem statements: their almost perfect performance on ordinal problems (94.79%) dropped 12.5 points lower on cardinal problems. Interestingly, in this experiment the participants were not made aware of any form of time constraint: they were explicitly told to take as long as they needed, contrary to the instructions given to the experts in Gros et al.'s study. Despite not being pressured by time, they still failed to identify the solution of a significant portion of cardinal problems. Of note, despite their lower performance on cardinal problems, their rate of success was comparable to that of the mathematical experts in this previous study, suggesting that their experience as math teachers helped them achieve a comparable – if not higher – level of expertise on the problems under scrutiny.

This result underlines the prevalence of these effect of world semantics on our ability to engage in arithmetic reasoning. Even among a population who, arguably, possesses a significant level of mathematical expertise, a certain degree of pedagogical content knowledge about mathematics, and who is accustomed to using word problems to teach students about abstract mathematical concept, detaching from the problem statements' context was not trivial. In fact, even when they managed to identify that the cardinal problems could be solved, both populations required more time to do so, which we attribute to the need to engage in a semantic recoding of their initial representation into a new one, congruent with the 1-step algorithm. Overall, these results indicate that despite their extensive understanding of the didactical complexities behind word problem solving, teachers (and more particularly

math teachers) are not immune to robust interferences between their non-mathematical world knowledge and their mathematical understanding of the situations depicted in a problem statement. In other words, knowing what to be wary of in a problem statement, and understanding that what matters is not the depicted context but the mathematical structure of the situation instead are not enough to systematically perceive the deep structure of the problems.

Experiment 2

While it seems clear in light of the first experiment's results that the difference between cardinal and ordinal problems runs deep enough to interfere even with math teachers' understanding of arithmetic word problems involving relatively elementary mathematical notions, the question remains as to what exactly this distinction between cardinal and ordinal problems entails in terms of solving process. How can we assess our hypothesis that the use of cardinal quantities fosters a set-based encoding enticing participants to calculate the value of each subset to find the value of the total set? While previous works have used algorithm choice and drawing production to provide evidence towards this claim (Gros et al., 2017), we strove to collect evidence from objective measurements that would not solely rest on participants verbal or written productions. To fill this gap, we decided to analyze the gaze patterns of 50 preservice teachers engaging in the solving of cardinal and ordinal problems similar to those used in Experiment 1. An additional measure of pupillometry was also designed to take a closer look at the claim that participants need to engage in a semantic recoding process to overcome a fruitless initial representation of the problem situation.

We made the following predictions. First, the difference in success rates observed in Experiment 1 between cardinal and ordinal problems should be replicated in this experiment. Second, the response times should also vary depending on the cardinal versus ordinal nature of the problems: correctly solving a cardinal problem should require a longer time than correctly solving an ordinal problem, as seen in the first experiment. Third, in accordance with the hypothesis that cardinal quantities evoke a set-based encoding fostering the use of a solving strategy consisting in calculating the values of *Part 2* and *Part 3* in order to get the value of *Whole 2*, and in accordance with the hypothesis that ordinal quantities evoke an axis-based encoding favoring the view that the value of *Whole 2* can be calculated by simply subtracting the value of the difference from the value of *Whole 1* (see Fig. 1), we

predicted that the total fixation time spent on each line of the problems should vary between cardinal and ordinal problems. Since a cardinal encoding is supposed to foster the calculation of the values of *Part 2* and *Part 3* to find the value of *Whole 2*, we expected that cardinal problems would lead to longer visit durations on the lines referring to *Part 2* and *Part 3*, compared to ordinal problems. Fourth, since the values of *Part 2* and *Part 3* are not provided in the problem statements, but are nevertheless deemed necessary by participants who encode the problems as a cardinal situations, then backward eye movements to the lines referring to these two quantities should be more frequent on cardinal than on ordinal problems. Fifth, since participants who manage to solve cardinal problems are supposed to do so by engaging in a costly semantic recoding process, then correctly solving a cardinal problem should result in an increase in pupil diameter whereas correctly solving an ordinal problem should not. Overall, these measures were meant to depict a more accurate picture of the interpretative mechanisms at play in the solving of cardinal and ordinal word problems.

Methods

Participants. The participants were 50 pre-service teachers (41 women and 9 men, $M = 27.22$ years, $SD = 13.95$) recruited from the Educational Sciences program at the University of Geneva. They volunteered after giving informed consent, in exchange for course credit. All the participants spoke French fluently and none had previously participated in any similar experiment. They were invited to assess the validity of 18 problems, during which their eye movements were recorded in a natural, dynamic, and undisturbed manner.

Materials. The arithmetic word problems used in this experiment were taken from the pool of 12 problems created in Gros et al. (2019). In order to maximize the number of recordings for each participant, we created 6 new problems to constitute a pool of 18 problems to choose from. Each participant was presented with a random selection of 12 target solvable problems: 6 of those featured cardinal quantities (2 collection problems, 2 price problems and 2 weight problems) and 6 featured ordinal quantities (2 duration problems, 2 height problems and 2 floor problems). We used a within-subject design to allow for within-subject comparisons between performance on cardinal and on ordinal problems. In addition, we introduced 6 unsolvable filler problems that were similar to the target problems but could not be solved with any algorithm, due to the missing value of *Whole 2* (see Fig. 1). Order of target problems

and filler problems was randomized between participants. All the problems were written in French. Cardinal and ordinal problems did not significantly differ in number of words; $t(18) = 1.92$, $p = .073$, independent t-test. The problems were isomorphs and the numerical values used were randomized across problems.

Procedure. The stimuli were presented on a 23.8" monitor. The participants were seated approximately 65 centimeters from the monitor, in a soundproofed experimental room designed to avoid distractions during the experiment. There was no window to avoid any natural light fluctuation; the only sources of light were a ceiling light and the eye tracking monitor. The first screen displayed the instructions for the experiment. Participants were told that "You will be presented with a series of arithmetic problems. Some of the problems can be solved using the values provided, while other problems cannot be solved with the available information. Your task is to tell apart problems that can be solved from problems that cannot. Answer as quickly as you can, although being correct is more important than being fast. Press the space bar when you are ready to start.". A fixation cross was displayed for 3 seconds before each problem appeared.

Each problem screen comprised 6 lines of text composing the problem statement, and a separate insert displaying the response choices. The text was written in size 18, with a line spacing of 3.7 to ensure that minor inaccuracies of the eye gaze estimation would not be detrimental to the analyses. The response insert presented two possible choices. Choice "A" was the solution to the problem, composed of the operation and a short sentence describing the result (e.g. "14 - 2 = 12. Lucy arrives at the 12th floor."). Choice "B" stated: "There is not enough information to find the solution." For each solvable target problem, the correct solution was proposed, and it was up to the participants to assess whether they thought it was valid or whether they deemed the problem unsolvable. Participants answered each problem using two keys on a keyboard placed in front of them. After each problem, a screen appeared informing them that their answer had been recorded and telling them to press the space bar when they were ready for the next problem. This procedure was the same for each problem. After completing the test, the participants were debriefed and dismissed. A typical session lasted between 20 and 30 minutes. The eye movements were registered with a Tobii Pro Spectrum eye tracker attached to the monitor.

Results

Success Rates. We recorded participants' answers to the problems to evaluate whether the results from Gros et al. (2019) and our first experiment were replicated. As in previous studies, we analyzed participants' rate of correct answers and response times (RTs) on the cardinal and ordinal target problems. The dependent variable was the failure or success on solvable problems (see Fig. 4). Since each participant gave a binary answer to 6 cardinal problems and 6 ordinal problems, we used a generalized linear mixed model with a binary distribution to account for the repeated measures in the experimental design. We used the cardinal versus ordinal nature of the problems as a fixed factor, and participants as a random effect. The overall model successfully converged and had a total explanatory power of 12.60% (conditional R^2). In line with previous results, participants performed significantly worse on cardinal (51.51%) than on ordinal problems (82.4%); $z = 8.22$, $p < .001$.

Response Times. Following our semantic recoding hypothesis, we looked at the RTs of correctly solved problems (see Fig. 5). We predicted that correctly solving a cardinal problem would require a higher amount of time, due to an extra recoding step being necessary to find the solution. We used Tukey's method to identify and remove 16 outliers ranged above and below 1.5 interquartile range. We analyzed participants' response times using a linear mixed model with the cardinal versus ordinal nature of the problems as a fixed factor and the participants as a random effect. The overall model successfully converged and explained 30.43% of the variance (conditional R^2). Within this model, an ANOVA using Satterthwaite's method for estimation of degrees of freedom revealed that there was a significant effect of the cardinal versus ordinal nature of the problems on the response times on correctly solved problems ($F(1) = 25.24$, $p < .001$). Which indicates that correctly solving a cardinal problem required more time on average ($M = 25.29$, $SD = 8.72$) than correctly solving an ordinal problem ($M = 21.73$, $SD = 7.73$). The results from Gros et al. (2019) were thus replicated in this experiment, both in terms of success rate and response times.

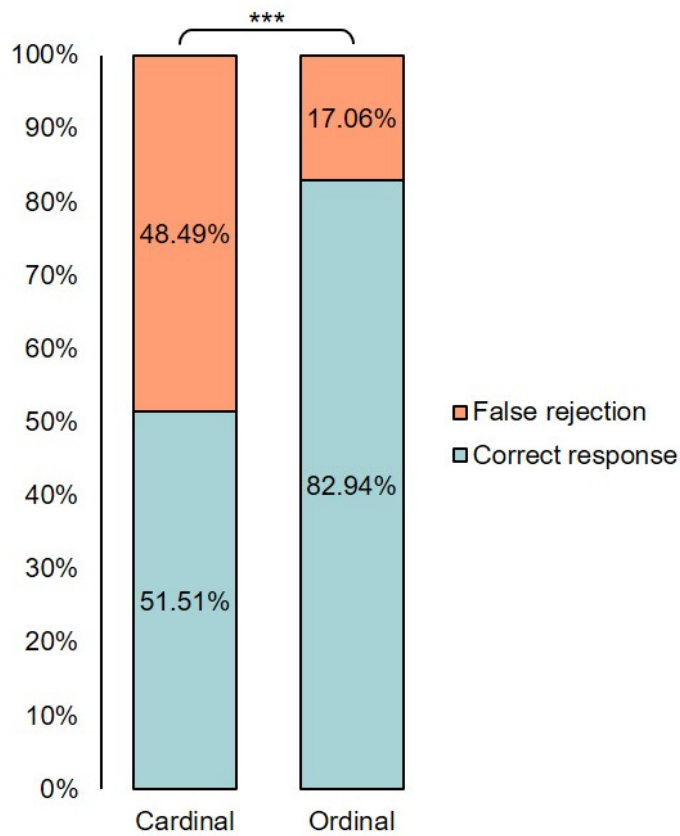


Fig. 4. Participants' answer distribution depending on the cardinal versus ordinal nature of the problems. *** $p < .001$.

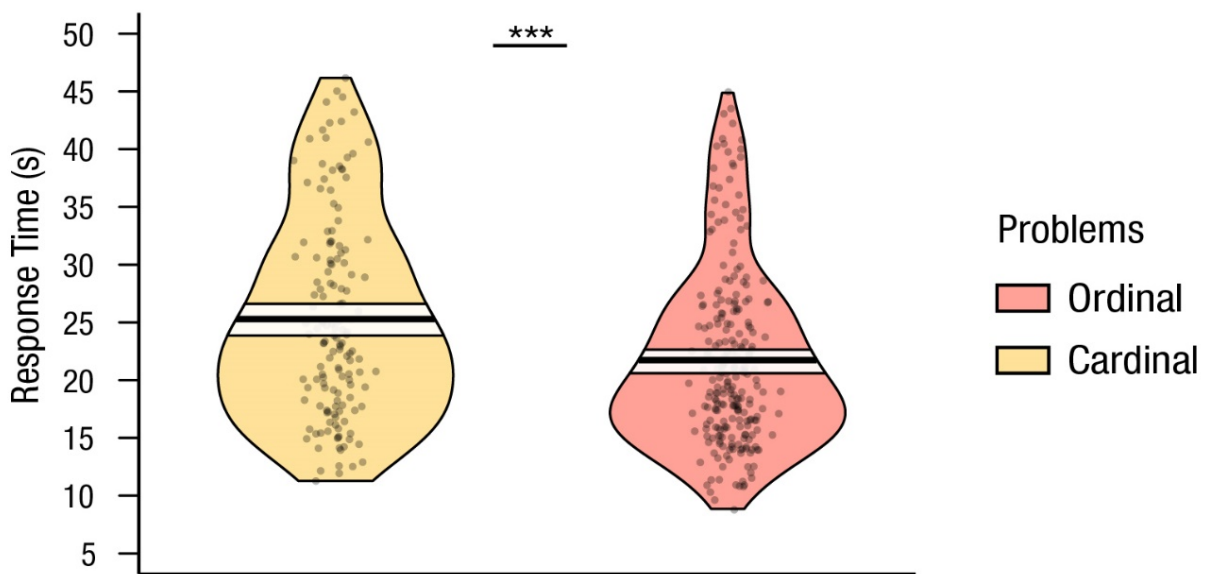


Fig. 5. Pirate plot of response times on cardinal and ordinal problems. Middle lines indicate mean RT, upper and lower lines indicate 95% confidence interval margins. *** $p < .001$.

Scoring of eye-fixation data. The sequence of eye fixations for each participant was recorded with the software Tobii Pro Lab, which identified the fixations and saccades of the participants while they were solving the problems. We partitioned the screen into 7 different areas of interest (AOIs) in which participants' fixations were recorded. In order to avoid false negative due to potential measurement inaccuracy, and since we had chosen the largest possible line spacing to avoid vertical overlapping between AOIs, we followed Orquin, Ashby, & Clarke's (2016) advice and selected the largest possible AOIs that were not susceptible to overlap. Thus, we created one AOI for each problem line, and one last AOI dedicated to the response insert. The seven AOIs partitioned the screen and they all shared the same height and width.

Looking time. An examination of the participants' fixations within the different AOIs provides a test of the hypothesis that participants attend to different pieces of information depending on whether they are solving a cardinal or an ordinal problem. We had predicted that, since cardinal problems are supposed to lead to a cardinal encoding of the situation, participants will spontaneously try to calculate the intermediate values of *Part 2* and *Part 3* to find the value of *Whole 2* (See Fig. 1, box 2.d.). Thus, participants should spend more time in sentences referring to these two subsets on cardinal problems than on ordinal problems. We were thus expecting an interaction between the cardinal versus ordinal nature of the problems and the total amount of time spent in the lines specifying the intermediate subsets *Part 2* (lines 2 and 4) and *Part 3* (line 5; see Fig. 1, box 2.b.).

We extracted the total visit duration per AOI for each participant. This measure included all the fixations within the AOI to indicate how long in total the participants had spent looking at each of the 6 problem lines as well as the response insert. Since each participant's gaze was recorded on 12 different problems, we analyzed the visit duration using a generalized linear mixed model with visit duration as the dependent factor, participants as a random effect, the line number as a fixed effect and the cardinal versus ordinal nature of the problems as a fixed effect. The overall model successfully converged and had a total explanatory power of 31.92% (conditional R²). Within this model, an ANOVA using Satterthwaite's approximation for the degrees of freedom revealed that the cardinal versus ordinal nature of the problems had a significant effect ($F(1) = 53.74, p < .001$), as well as the line number ($F(6) = 202.69, p < .001$). The interaction between those two fixed effect was significant as well ($F(6)$

= 12.29, $p < .001$). In accordance with our hypothesis, we computed orthogonal contrasts using least square means to identify whether participants did visit the lines referring to *Part 2* and *Part 3* longer on cardinal than on ordinal problems: lines 2, 4 and 5. Results revealed that participants spent a longer time on average visiting line 2 on cardinal problems ($M = 3.40$ seconds, $SD = 1.64$) than they did on ordinal problems ($M = 2.58$ seconds, $SD = 1.43$); $t(3922) = 6.49$, $p < .001$. Similarly, participants spent a longer time visiting line 4 on cardinal problems ($M = 4.40$ seconds, $SD = 1.66$) than they did on ordinal problems ($M = 3.61$ seconds, $SD = 1.60$); $t(3924) = 6.13$, $p < .001$. Finally, they also spent a longer time on the 5th line of cardinal problems ($M = 3.79$ seconds, $SD = 1.75$) than on that of the ordinal problems ($M = 2.97$ seconds, $SD = 1.67$); $t(3923) = 6.56$, $p < .001$. On the other hand, there was no significant visit duration difference between cardinal and ordinal problems on lines 1, 3, 6 nor on the response insert (see Fig. 6); $0.02 \leq t\text{-value} \leq 0.96$, $.33 \leq p \leq .98$.

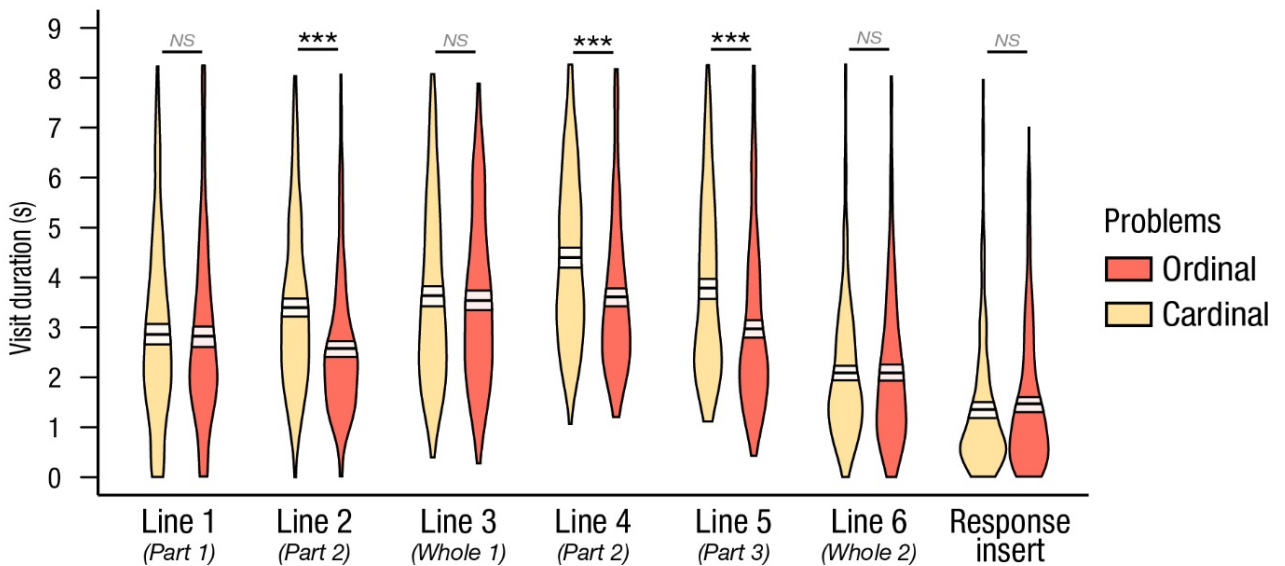


Fig. 6 Visit duration per problem line

Due to two thirds of the problems we used coming from a previous study (Gros et al., 2019), we could not perfectly control for the word length of every line in the problems. While there was no length difference in lines 2 to 6, there was however a significantly higher number of words in line 1 of ordinal problems ($M = 6.89$, $SD = 2.15$) as compared to line 1 of cardinal problems ($M = 10.22$, $SD = 2.82$); $t(16) = 2.82$, $p < .05$. This difference was not deemed problematic regarding our results since our hypotheses focused on lines 2, 4 and 5, and since there was no significant difference of the visit duration on line 1 between cardinal and ordinal problems.

Regressions. Analysis of backward eye movements (regressions) from one problem sentence to an earlier one is another measure of participants' use of the pieces of information available in the problem statement. Following Verschaffel, De Corte and Pauwels' advice (1992), we investigated participants' number of regressions to get a finer-grained understanding of the differences between cardinal and ordinal problems. Since each problem line presented a new information, we could infer which pieces of information participants were going back to in order to try to solve the problems. For each trial, we calculated the total number of backward eye movements to lines to each problem line. We started counting the backward eye movements only after the participants had made their first fixation on the first line of the problems. As in the visit duration analysis, we predicted that since cardinal problems are supposed to evoke a cardinal encoding fostering the calculation of *Part 2* and *Part 3* to get the value of *Whole 2* (see Fig. 1, box 2.d.), then participants should make more regressions to the lines mentioning these quantities in their search for the missing values needed to use the 3-step algorithm. Thus, we made the hypothesis that participants would make a higher number of regressions to lines 2, 4 and 5 on cardinal problems than they would on ordinal problems.

We performed the analysis using a generalized linear mixed model with number of regressions as the dependent factor, the cardinal versus ordinal nature of the problems as a fixed factor, the line number as a fixed factor and the participants as a random effect. The overall model successfully converged and had a total explanatory power of 24.75% (conditional R^2). Within this model, an ANOVA using Satterthwaite's method for estimation of the degrees of freedom revealed that the effect of the cardinal versus ordinal nature of the problems was statistically significant ($F(1) = 140.11, p < .001$). There was also a main effect of the line number ($F(5) = 94.43, p < .001$). The interaction between these two fixed factors was significant ($F(5) = 16.36, p < .001$). In accordance with our hypothesis, we computed orthogonal contrasts using least square means to identify whether participants did make more regressions to lines 2, 3 and 5 on cardinal problems than they did on ordinal problems (see Fig. 7).

Results revealed that, as predicted, participants made a higher number of regressions to line 2 on cardinal problems ($M = 2.03$ regressions per problem, $SD = 1.73$) than they did on ordinal problems ($M = 1.05$ regressions per problem, $SD = 1.16$); $t(3306) = 9.39, p < .001$. Similarly, participants made more regressions to line 4

on cardinal problems ($M = 1.23$, $SD = 1.43$) than on ordinal problems ($M = 0.70$, $SD = 1.01$); $t(3306) = 5.12$, $p < .001$. Finally, the number of regressions to line 5 was higher on cardinal problems ($M = 0.77$, $SD = 1.11$) than it was on ordinal problems ($M = 0.42$, $SD = 0.76$); $t(3306) = 3.43$, $p < .001$. Surprisingly, the contrast analysis also revealed a difference that we had not predicted: participants made a higher number of regressions per problem to line 1 on cardinal problems ($M = 2.00$, $SD = 1.83$) than they did on ordinal problems ($M = 1.01$, $SD = 1.24$); $t(3306) = 9.63$, $p < .001$. There was no such difference between cardinal and ordinal problems on line 3 ($t(3306) = 1.75$, $p = .08$) nor on line 6 ($t(3306) = 0.31$, $p = .76$).

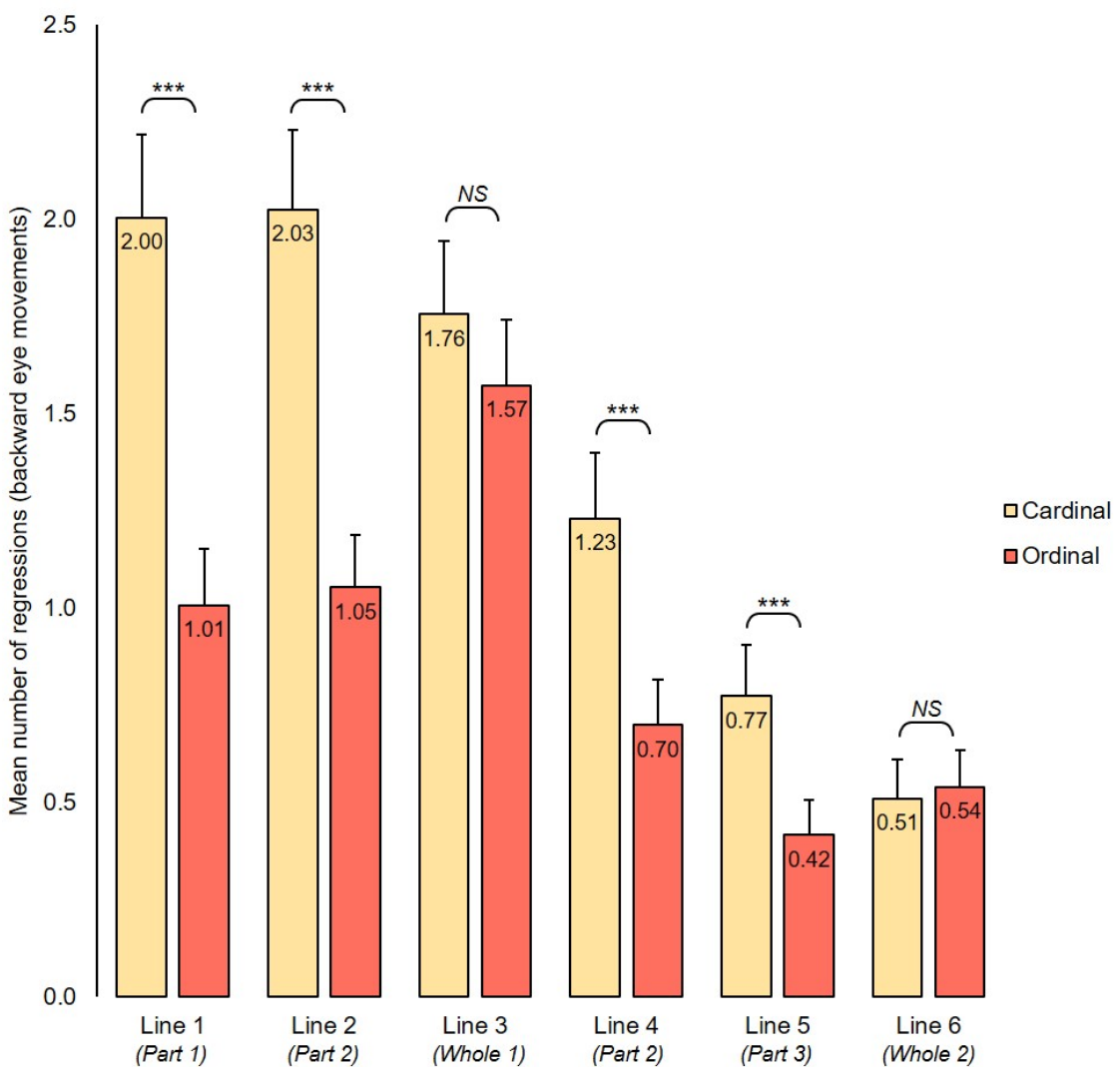


Fig. 7 Mean number of regressions to specific lines depending on the nature of the problems. Error bars indicate upper margins of 95% confidence intervals.

Pupillary dilatation. To evaluate the validity of our claim that participants need to semantically recode their initial representation of cardinal problems to find the solution, we looked at participants' pupil dilation in relation with their successes and failures in solving the problems. Since task-evoked pupillary response is deemed a reliable measure of cognitive load (van der Wel & van Steenbergen, 2018), the analysis of inter-trial change in pupil diameter should be informative with regards to the existence of a semantic recoding process. We measured participants' pupil diameter on each fixation during each problem and contrasted it with their answers to the problems. We analyzed inter-trial change in pupil diameter using a generalized linear mixed model with pupil diameter during fixations as the dependent variable. We used participants as a random effect, the cardinal versus ordinal nature of the problems as a fixed factor and the participants' response to the problems as a fixed factor. The overall model successfully converged and explained 85.21% of the variance (conditional R²). Within this model, an ANOVA using Satterthwaite's approximation for the degrees of freedom revealed that the cardinal versus ordinal nature of the problems had a significant effect ($F(1) = 68.87, p < .001$), indicating that pupil dilation differed between cardinal and ordinal problems. There was no main effect of the response provided by the participants ($F(1) = 0.38, p = .54$). There was however an interaction between the type of problem (cardinal/ordinal) and the response given by the participants (true/false): $F(1) = 5.73, p < .05$.

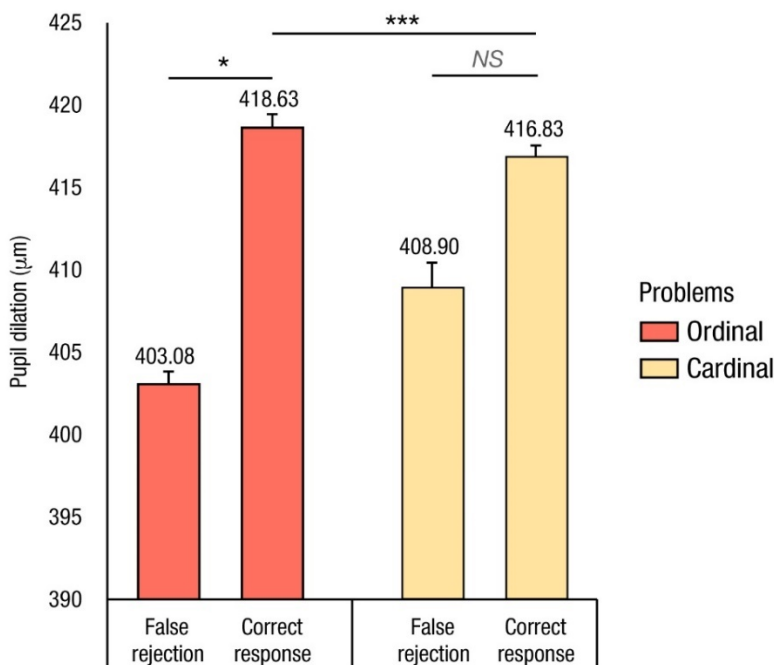


Fig. 8 Pupil dilation on solvable problems. Error bars indicate upper margins of 95% confidence intervals.

In accordance with our hypothesis that semantic recoding needs to happen on cardinal problems but not on ordinal problems, we computed contrasts using least square means to identify whether participants' response was linked to their pupil dilation on cardinal problems and on ordinal problems (see Fig. 8). Results indicated that correctly solving a cardinal problem was associated with a larger pupil diameter on average as compared to correctly solving an ordinal problem ($t(58838) = 5.34, p < .001$), which suggests that finding the solution to cardinal problems was more cognitively taxing than finding the solution to ordinal problems. Besides, a comparison of successes and failures revealed that participants' pupil diameter was significantly larger on correctly solved cardinal problems ($M = 418.63 \mu\text{m}, SD = 53.14$) than on failed cardinal problems ($M = 403.08 \mu\text{m}, SD = 49.05$); $t(58841^3) = 2.48, p < .05$. On the other hand, there was no significant difference between pupil diameter on correctly solved ordinal problems ($M = 416.83 \mu\text{m}, SD = 54.18$) and on incorrectly rejected ordinal problems ($M = 408.90 \mu\text{m}, SD = 55.88$); $t(58841) = 1.09, p = .28$. Thus, there was evidence of an increase in cognitive load on cardinal problems correctly solved, but no such evidence on ordinal problems.

Discussion

In this second experiment, we gathered converging evidence from five different sources of information regarding the interpretative processes at play in the encoding, recoding and solving of cardinal and ordinal problems. First, the success rate analysis confirmed the results of the first experiment regarding the increased difficulty to perceive the validity of the 1-step algorithm on cardinal problems as compared to ordinal problems. Then, the difference in response times between correctly solved cardinal and ordinal problems was also replicated, supporting the hypothesis that one needs to engage in a semantic recoding step to construct a new representation compatible with the 1-step algorithm.

Third, by studying the total looking time on each line of the problem using eye-tracking technology, we were able to take a closer look at what differentiates the encoding of cardinal and ordinal problems. We hypothesized that problems using cardinal quantities would lead participants to abstract a cardinal encoding of the situation emphasizing the set/subset structure of the situation depicted. Thus, in their

³ The high number of degrees of freedom for this analysis is due to pupil dilation being measured at each time step for each fixation recorded by the eye-tracker. We are currently looking into the possibility to perform a different analysis focusing on participants' pupil dilation variations compared to a baseline.

attempts to find the value of *Whole 2*, we predicted that participants' first reaction would be to try to find the values of each of its subsets, that is, *Part 2* and *Part 3* (see Fig. 1). Our looking time analysis revealed that it was indeed the case, since lines 2, 4 and 5 were visited for a longer time on cardinal problems than on ordinal problems. Despite the lines presenting the same information in the same order across problems, it seems that these three specific lines received particular attention on cardinal problems, thus suggesting that cardinal problems emphasized the importance of *Part 2* and *Part 3* to find the solution. This results supports the idea that cardinal quantities evoke a set-based representation. On the other hand, the lines presenting pieces of information that were useful to solve the problems but that were not directly related to *Part 2* or *Part 3* were not looked at significantly longer on cardinal than on ordinal problems.

Fourth, the analysis of backward eye movements from one line to a previous one informed us with regards to the information that participants came back to when reading the problems. In accordance with the results of the visit duration analysis, participants made more regressions to lines 4 (*Part 3*) as well as to lines 2 and 5 (*Part 2*) on cardinal problems, as compared to ordinal problems. This indicates that participants' strategy includes looking back to previous lines for information about *Part 2* and *Part 3*, thus supporting the idea that participants were actively trying to find the value of *Whole 2* by adding up the (missing) values of *Part 2* and *Part 3*. This result is in accordance with the visit duration analysis, although it provides a different perspective on the matter: the visit duration analysis told us that *Part 2* and *Part 3* were the focus of special attention on cardinal problems, whereas the regression analysis told us that participants tended to look back at information regarding *Part 2* and *Part 3* more often on cardinal problems. In other words, the pre-service teachers in our study reinspected the information regarding *Part 2* and *Part 3* more often on cardinal problems, which backs the idea that their initial encoding of the cardinal problems lead them to look for values that were not provided in the statement nor could they be calculated. An unexpected effect appeared during the contrast analysis regarding the regressions made toward line 1: participants tended to reinspect line 1 more often on cardinal than on ordinal problems. This means that in addition to *Part 2* and *Part 3*, participants also tended to come back to *Part 1* in their attempts to solve the problem. While this result was not within the range of our predictions, we can still attempt to provide a post-hoc explanation for its existence. Line 1 introduces the

value of *Part 1*, and since the value of *Part 3* is introduced in line 5 by comparison to *Part 1* (e.g. “The German dictionary (*Part 3*) is 2 kgs lighter than the Russian dictionary (*Part 1*).”), then calculating the value of *Part 3* would require knowing the value of *Part 1*. In other words, a possible explanation for this difference would be that in trying to figure out the value of *Part 2*, participants would be led to re-evaluate the first sentence introducing *Part 1*, since the value of *Part 2* is always given relatively to the value of *Part 1*. Further investigation is needed on this question to identify the proper explanation for this difference.

Finally, a fifth measure provided new insights regarding our hypothesis that solving a problem whose initial representation is semantically incongruent with its solving algorithm requires to engage in a costly semantic recoding process to construct a new representation compatible with the available algorithm. Pupil dilation varies closely in response to changes in task demands and pupillometry can thus be used as an indirect measure of participants’ effort (van der Wel & van Steenbergen, 2018). By studying pupil dilation variations between success and failures on cardinal and on ordinal problems, we were able to measure how the cognitive load varied between situations. In accordance with our prediction regarding response times, we predicted that participants’ engagement in a semantic recoding step would result in an increase in pupil diameter on successfully solved cardinal problems. The results supported this hypothesis, since there was an increase in pupil diameter on successfully solved cardinal problems as compared to erroneously rejected cardinal problems. In other words, pupil dilation indicated an increased effort when participants managed to overcome their initial, incongruent representation of the problems and to find the solution to the cardinal problems. On the other hand, the pupil diameter difference between successes and failures on ordinal problems was not statistically significant. This can either indicate that there was no such difference since no semantic recoding was needed on ordinal problems, or it can simply be the sign of a lack of statistical power, since failures on ordinal problems were relatively scarce. Although we do not have the means to arbitrate between these two candidate explanations, the fact that there remained a significant difference between pupil dilation on correctly solved cardinal problems and on correctly solved ordinal problems seems to tip the scale in favor of the first interpretation. Indeed, it appears that correctly solving a cardinal problem required more effort, on average, than

correctly solving an ordinal problem, which could be a sign of the existence of the semantic recoding process we hypothesized.

General discussion

When it was made clear, in Gros et al. (2019), that even mathematicians' expertise could be thwarted by the use of dictionaries instead of ages in a problem statement whose answer was a single-step subtraction, we came to the realization that there was only two possible explanations to this result. Either the influence of world knowledge on the encoding and solving of arithmetic word problems was so strong that no kind of expertise could be immune to it; which would suggest that humans are irremediably incapable to spontaneously detach their mathematical reasoning from the context in which they apply it. Or it could also be that the idiosyncratic nature of arithmetic word problems placed them outside of the scope of mathematicians' expertise. In other words, despite being experts of abstract, context-independent reasoning (Dehaene, 2011), mathematicians might have been too unaccustomed to the peculiar form or arithmetic word problems to be able to optimally tap into their mathematical expertise. In an attempt to answer this question, we decided in this study to recruit participants enjoying an expertise both in arithmetic, and in its teaching in schools, by the means of arithmetic word problems. To put it in Shulman's (1986) terms, they possessed both Content Knowledge and Pedagogical Content Knowledge. However, results from our first experiment showed that even that was not enough to systematically perceive that cardinal problems could be solved with a single-step subtraction, just like ordinal problems. Teachers' irrelevant, non-mathematical knowledge about the world interfered with their ability to identify the solution to the word problems, despite being given as long as they needed to solve the problems.

The importance of this finding lays in the central role that teachers' diagnostic judgments hold in student-centered teaching approaches (Davis & Simmt, 2006; Ostermann, Leuders, & Nückles, 2018; Prediger & Zindel, 2017). Indeed, teachers' knowledge of students' conceptions and misconceptions is considered an essential component of their Pedagogical Content Knowledge, along with their knowledge of instructional strategies and representations (Ball, Thames, & Phelps, 2008; Depaepe, Verschaffel, & Kelchtermans, 2013). Thus, identifying the difficulties that students

might encounter and understanding the conceptions underlying them is a crucial step for teachers of any domain. Here, by showing that teachers failed to solve problems that were framed in a context semantically incongruent with their solution, our results suggest that teachers are not immune to the difficulties their students are met with. This finding is in direct line with a recent study showing that teachers' intuitive conceptions may overshadow their pedagogical content knowledge, causing them to struggle with assessing the difficulty of arithmetic word problems (Gvozdic & Sander, 2018).

In light of this observation, we strove to get a finer understanding of what exactly happens when one's non-mathematical knowledge interferes with one's mathematical expertise. The second experiment brought new evidence towards the claim that the use of cardinal versus ordinal quantities has a direct influence on the problem representations that are constructed by the solvers, as well as on their ability to find the solution. By showing an increased focus on the subsets on cardinal problems, the use of eye-tracking helped support the idea that set-based representations are constructed whenever weights, prices or collections are mentioned. Additionally, pupil dilation analysis confirmed that recoding an initially incongruent representation is a demanding process, with an important cognitive cost associated to it. This is an important step forward in understanding how to foster semantic recoding in schools, which is a crucial prerequisite to generate transfer between superficially dissimilar situations sharing a deeper bond.

References

- Andrá, C., Lindström, P., Arzarello, F., Holmqvist, K., Robutti, O., & Sabena, C. (2015). Reading mathematics representations: an eye-tracking study. *International Journal of Science and Mathematics Education*, 13(2), 237–259.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special. *Journal of teacher education*, 59(5), 389–407.
- Bassok, M. (2001). Semantic alignments in mathematical word problems. In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 401–433). Cambridge, MA: MIT Press.
- Beatty, J., Lucero-Wagoner B (2000) The pupillary system. In: Cacioppo JT, Tassinari LG, Berntson G, (Eds.), *Handbook of Psychophysiology*. (pp. 42–162.). Cambridge, UK: Cambridge University Press.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, 91(2), 276.
- Beitlich, J., Obersteiner, A., & Reiss, K. (2015). How do secondary school students make use of different representation formats in heuristic worked examples? An analysis of eye movements. *Proceedings of PME 39*, 2, 97-104.
- Bermejo, V. (1996). Cardinality development and counting. *Developmental Psychology*, 32(2), 263-268.
- Bolden, D., Barmby, P., Raine, S., & Gardner, M. (2015). How young children view mathematical representations: a study using eye-tracking technology. *Educational research*, 57(1), 59-79.
- Buswell, G. T. (1935). *How people look at pictures: a study of the psychology and perception in art*. Chicago, IL: The University of Chicago Press.
- Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain*. Amsterdam, Netherlands: Elsevier. 341-371.
- Colomé, À., & 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.

- Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, *137*(1), 22-38.
- Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive psychology*, *20*(4), 405-438.
- Curtis, E. T., Huebner, M. G., & LeFevre, J. A. (2016). The relationship between problem size and fixation patterns during addition, subtraction, multiplication, and division. *Journal of numerical cognition*, *2*(2), 91-115.
- 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.
- Davis, B., & Simmt, E. (2006). Mathematics-for-teaching: An ongoing investigation of the mathematics that teachers (need to) know. *Educational studies in mathematics*, *61*(3), 293-319.
- De Corte, E., & Verschaffel, L. (1986). Eye-Movement Data as Access to Solution Processes of Elementary Addition and Subtraction Problems. *Annual Meeting of the American Educational Research Association*, San Francisco, April 16-20, 1986.
- De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of Educational Psychology*, *82*(2), 359-365.
- Dehaene, S. (2011). *The number sense: How the mind creates mathematics* (Rev. and updated ed.). New York, NY: Oxford University Press.
- Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and teacher education*, *34*, 12-25.
- Dewolf, T., Van Dooren, W., Hermens, F., & Verschaffel, L. (2015). Do students attend to representational illustrations of non-standard mathematical word problems, and, if so, how helpful are they?. *Instructional Science*, *43*(1), 147-171.
- Fayol, M., Barrouillet, P., & Camos, V. (1997). Early mathematics learning: What can research tell us. In *Communauté Européenne, Convention DG*, *22*, 96-101.
- Fischer, F. E., & Beckey, R. D. (1990). Beginning kindergarteners' perception of number. *Perceptual and Motor Skills*, *70*(2), 419-425.
- Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.

- 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.
- Gelman, R., & Gallistel, C. R. (1978). *The child's concept of number*. Cambridge, MA: Harvard University Press.
- Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science, 21*(2), 90-95.
- Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33*(4), 457-461.
- Green, H. J., Lemaire, P., & Dufau, S. (2007). Eye movement correlates of younger and older adults' strategies for complex addition. *Acta psychologica, 125*(3), 257-278.
- 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*, online first.
- Gros, H., Thibaut, J. P., & Sander, E. (2017). The nature of quantities influences the representation of arithmetic problems: Evidence from drawings and solving procedures in children and adults. In R. Granger, U. Hahn, & R. Sutton (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp 439-444). Austin, TX: Cognitive Science Society.
- Gvozdic, K., & Sander, E. (2018). When intuitive conceptions overshadow pedagogical content knowledge: Teachers' conceptions of students' arithmetic word problem solving strategies. *Educational Studies in Mathematics, 98*(2), 157-175.
- Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology, 84*(1), 76-84.
- Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of educational psychology, 87*(1), 18-32.
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*(3611), 1190-1192.
- Hyönä, J., Lorch, R. F. Jr, & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach & H. Deubel (eds.), *The mind's eye:*

- Cognitive and applied aspects of eye movement research*. Amsterdam, North Holland: Elsevier Science. 313-334.
- Janisse, M. P. (1977). *Pupillometry: The psychology of the pupillary response*. New York, NY : Halsted Press.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Cambridge, MA: Harvard University Press.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4), 329-354.
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Needham Heights, MA: Allyn & Bacon.
- Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, 4(1-2), 56-88.
- Kaakinen, J. K., & Hyönä, J. (2008). Perspective-driven text comprehension. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22(3), 319-334.
- Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154(3756), 1583-1585.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological review*, 92(1), 109-129.
- Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & cognition*, 29(7), 1000-1009.
- Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., ... & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10, 90-115.
- Latour, P. L. (1962). Visual threshold during eye movements. *Vision Research*, 2(3), 261-262.
- 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.

- Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational psychology*, *79*(4), 363-371.
- Merkley, R., & Ansari, D. (2010). Using eye tracking to study numerical cognition: the case of the ratio effect. *Experimental Brain Research*, *206*(4), 455-460.
- Meyer, C., Barbiers, S., & Weerman, F. (2016). Order and Ordinality: The Acquisition of Cardinals and Ordinals in Dutch. In *Boston University Conference on Language Development (BUCLD) 40*. Cascadilla.
- 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), 129-140.
- 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.
- Nesher, P., & Teubal, E. (1975). Verbal cues as an interfering factor in verbal problem solving. *Educational Studies in Mathematics*, *6*(1), 41-51.
- Obersteiner, A., & Tumpek, C. (2016). Measuring fraction comparison strategies with eye-tracking. *ZDM*, *48*(3), 255-266.
- Orquin, J. L., Ashby, N. J., & Clarke, A. D. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making*, *29*(2-3), 103-115.
- Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta psychologica*, *144*(1), 190-206.
- Ostermann, A., Leuders, T., & Nückles, M. (2018). Improving the judgment of task difficulties: Prospective teachers' diagnostic competence in the area of functions and graphs. *Journal of Mathematics Teacher Education*, *21*(6), 579-605.
- Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology*, *11*(5), 559-566.
- Pollak, H. O. (1969). How can we teach applications of Mathematics?. *Educational studies in mathematics*, *2*(2), 393-404.
- Prediger, S., & Zindel, C. (2017). Deepening Prospective Mathematics Teachers' Diagnostic Judgments: Interplay of Videos, Focus Questions and Didactic Categories. *European Journal of Science and Mathematics Education*, *5*(3), 222-242.

- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372-422.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10, 241–255.
- Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements?. *Vision research*, 16(8), 829-837.
- Riley, M. S., Greeno, J. G., & Heller, J. (1983). Development of children's problem-solving ability in arithmetic. *The development of mathematical thinking*, 153-196.
- Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of experimental child psychology*, 103(3), 325-337.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans and understanding: An inquiry into human knowledge structures*. New York, NY: Hillsdale.
- Scheiter, K., & van Gog, T. (2009). Introduction: Using eye tracking in applied research to study and stimulate the processing of information from multi-representational sources. *Applied Cognitive Psychology*, 23(9), 1209-1214.
- Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smedt, B., Verschaffel, L., ... & Stern, E. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development*, 23(3), 409-422.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, 15(2), 4-14.
- Staub, F. C., & Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In C. A. Weaver III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 285-305). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Stern, D. (1993). L'enveloppe prénarrative. *Journal de la psychanalyse de l'enfant*, 14, 13-65.
- Strohmaier, A. R., Lehner, M. C., Beitlich, J. T., & Reiss, K. M. (*in press*). Eye movements during mathematical word problem solving – global measures and individual differences. *Journal für Mathematik-Didaktik*, 1-33.
- Strohmaier, A. R., Schiepe-Tiska, A., Chang, Y. P., Müller, F., Lin, F. L., & Reiss, K. M. (*in press*). Comparing eye movements during mathematical word problem solving in Chinese and German. *ZDM*, 1-14.

- Strohmaier, A. R., Tatsidou, K., Reiss, K. (2018). Eye movements during the reading of word problems. *Advances in the use of eye tracking data. Fachgruppe Didaktik der Mathematik der Universität Paderborn (Hrsg.), Beiträge zum Mathematikunterricht 2018* (S. 1759-1762). Münster: WTM-Verlag.
- Thevenot, C. (2017). Arithmetic Word Problem Solving: The Role of Prior Knowledge. In Geary, D. C., Berch, D. B., Ochsendorf, J., Mann-Koepke, K. (Eds.), *Acquisition of Complex Arithmetic Skills and Higher-Order Mathematics Concepts* (pp. 47-66). Academic Press.
- Thevenot, C., & Barrouillet, P. (2015). Arithmetic word problem solving and mental representations. *The Oxford handbook of numerical cognition*, 158-179.
- Van der Schoot, M., Bakker Arkema, A. H., Horsley, T. M., & van Lieshout, E. C. (2009). The consistency effect depends on markedness in less successful but not successful problem solvers: An eye movement study in primary school children. *Contemporary Educational Psychology*, 34(1), 58-66.
- Van der Schoot, M., Vasbinder, A. L., Horsley, T. M., Reijntjes, A., & van Lieshout, E. C. (2009). Lexical ambiguity resolution in good and poor comprehenders: An eye fixation and self-paced reading study in primary school children. *Journal of Educational Psychology*, 101(1), 21-36.
- van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic bulletin & review*, 25(6), 2005-2015.
- van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, 20(2), 95-99.
- Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology*, 84(1), 85-94.
- Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse: Swets & Zeitlinger.
- 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.
- Winoto, P., Tang, T. Y., Huang, Z., & Chen, P. (2017, July). "Thinking in Pictures?" Performance of Chinese Children with Autism on Math Learning Through Eye-

Tracking Technology. In *International Conference on Learning and Collaboration Technologies*. Springer, Cham.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358(6389), 749-750.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision* (pp. 171-211). Springer, Boston, MA.

Zhu, R., Luo, Y., You, X., & Wang, Z. (2018). Spatial bias induced by simple addition and subtraction: From eye movement evidence. *Perception*, 47(2), 143-157.

CHAPTER 6 – CAN CHILDREN THINK ‘STRAIGHT’?
ORDINALITY AND CARDINALITY IN ARITHMETIC
WORD PROBLEM SOLVING



Presentation

In an attempt to establish the pervasiveness of the effects of semantic (in)congruence, Chapters 2 to 5 have focused on adults' reasoning about arithmetic word problems. However, children are believed to rely more on informal knowledge to solve math problems than adults do. Chapter 6 thus aims at evaluating the influence that the use of cardinal and ordinal quantities has on children's apprehension of numerical situations, while at the same time addressing the developmental aspect of this question. Indeed, the development of the understanding of cardinality and ordinality in the early years of life has been the subject of extensive work. However, these studies have mainly focused on the first stages of learning how to count, and the influence that the distinction between cardinality and ordinality still holds once basic counting procedures are mastered has hardly been scrutinized. Are children influenced by their perception of ordinality and cardinality in their apprehension of numerical situations? Does this distinction result in different mental representations? How early can children construct ordinal representations of events? Can children construct set-based representations with the same efficiency as adults? We designed two experiments to investigate these questions.

In Experiment 1, we ask 59 fifth graders and 52 adults to solve the same problems and to make, for each of them, a drawing that could help someone else find the solution. This drawing task was introduced to provide us with a new source of information regarding the structure of the representations constructed while solving the problems. We predict that among both populations, the difference between cardinal and ordinal problems will influence their algorithm use as well as drawing productions: drawings of cardinal problems should involve a higher proportion of cardinal features whereas drawings of ordinal problems should involve a higher proportion of ordinal features. In Experiment 2, we study the influence of the cardinal versus ordinal distinction among an even younger population. We ask 384 second graders to try to solve a simplified version of the problems used in Experiment 1. Although they were modified to be solvable by children of 7-8 years old, the problems still admitted two distinct solving algorithms. We predict that even at this young age, differences in algorithm use will be observed between cardinal and ordinal problems.

Experiment 1 has been reported in an article in the *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (2017). The present chapter is an extension of this article, with a new experiment conducted among 2nd graders.

Abstract

Previous studies have showed that children progressively acquire the numerical principles of ordinality and cardinality when they first learn how to count. However, little work has been conducted on the importance that cardinality and ordinality still hold in later years in more complex arithmetic reasoning. We investigate children's and adults' representation of arithmetic word problems designed to promote either a cardinal or an ordinal encoding by referring to certain quantities (durations, heights, and number of floors vs. weights, prices, and collections). In Experiment 1, we ask 59 children and 52 adults to solve the same arithmetic problems, and to make drawings of the depicted situations. We show that the use of quantities emphasizing the cardinal (resp. ordinal) aspect of the problems' values results in drawings with higher cardinal (resp; ordinal) features, as well as in the use of different solving algorithms. In Experiment 2, we investigate how pervasive the effect of the cardinal-ordinal distinction on algorithm choice is, by asking 191 lay 2nd graders, 193 trained 2nd graders and 67 adults to solve isomorphic problems created to favor either a cardinal or an ordinal representation. We show that regardless of their age and arithmetic proficiency, participants' strategy use is influenced by the cardinal-ordinal distinction manipulated through the use of specific types of quantities.

Keywords

cardinality · ordinality · encoding · drawing · problem solving · arithmetic word problems

Introduction

Word problem solving is a central component of mathematics education. 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 works suggests that problems are encoded into a mental representation including both mathematical and non-mathematical information, which is then translated into a solving algorithm (Bassok, 2001; Gamo, Sander, & Richard, 2010; Gros, Sander, & Thibaut, 2019). While the traditional schema theory (Kintsch & Greeno, 1985) posits that learning how to solve mathematical word problems is tantamount to acquiring and storing new problem schemas, we argue instead that it implies learning how to construct semantic representations including information regarding the non-mathematical content of a problem. In this paper, we attempt to provide evidence for the development of such representations throughout the years by studying the differences between cardinal and ordinal encodings of mathematical word problems in children and adults.

Importance of content effects throughout development and education

We have known ever since Riley, Greeno and Heller's (1983) work on the typology of additive word problems that different wordings of otherwise mathematically similar problems may be associated with different performances. However, the idea that the semantic content of a problem statement itself might interfere with its interpretation and subsequent solving is rather recent in the mathematical problem solving literature.

One of the first studies to demonstrate the effect that an individual's prior knowledge may have on their mathematical word problem solving performance was conducted by Bassok, Wu and Olseth (1995), using permutation problems with adult participants. The authors showed that the semantic relations connecting a problem's entities influence analogical transfer between isomorphic problems. They notably contrasted problems where objects were given to people and problems where people were assigned to objects. They found that, since in real-life objects are usually given

to people rather than people being assigned to objects, objects-to-people training examples led to higher performances with objects-to-people transfer problems (89%) than with people-to-objects transfer problems (0%). Furthermore, when the training problems involved situations that were not compatible with participants' experience of the world (in the quoted research, when the problems involved the assignation of people to objects), then participants' performance dropped in the transfer task, regardless of the type of transfer problems being used. Along this line of work, Bassok, Chase and Martin (1998) asked adults to create addition or division problems involving specific sets of objects that were provided. They showed that when the objects shared a functionally asymmetric semantic relation (e.g. apples and baskets evoke the contain relation), participants tended to create division problems, whereas they created addition problems when using symmetric sets of objects (such as oranges and apples, that belong to the same superordinate fruit category). These biases are not driven by arithmetic properties but rather by the semantic relations existing between the entities mentioned.

To account for such effects, Bassok (2001) proposed that the solvers abstract an interpreted structure that depends on their world knowledge about the entities described in the problem statement. This interpreted structure integrates the structural role of the entities mentioned in the problem and can thus lead to an appropriate use of abstract formal knowledge when the relations it describes are semantically aligned with the mathematical relations of the problem (Bassok et al., 1998; Bassok, 2001). Both behavioral (Bassok, Pedigo, & Oskarsson, 2008) and physiological (Guthormsen et al., 2016) measures confirmed that arithmetic reasoning is easier when daily-life knowledge (world semantics) and knowledge about mathematical concepts (mathematical semantics) are aligned with each other.

Investigation of mental representations

As previously discussed, the idea that different representations are abstracted depending on the semantic content of a problem statement is a promising one, as it provides an account of some performance differences reported in the literature. However, since direct investigation of mental constructs is seldom possible, evaluating the existence as well as the nature of these representations necessarily requires taking an indirect route. Numerous measures of the content of an individual's representation have been proposed, with varying degrees of success, from the inherently biased introspection (e.g. Wundt, 1907) to more objective – although harder to interpret –

physiological measures, such as EEG (Berkum, Hagoort, & Brown, 1999) or fMRI (*e.g.* Ischebeck, Schocke, & Delazer, 2009). Intermediate sources of information into one's representations have also been identified, such as metaphors (*e.g.* Lakoff & Nuñez, 2000), gestures (*e.g.* Fuhrman & Boroditsky, 2010), or written statements (*e.g.* Pinnegar, Mangelson, Reed, & Groves, 2011). In this study, we decided to bring together two complementary methods to study participants' interpretative processes: drawing production and algorithm choice.

Drawing production appears to be a promising path to investigate one's representation without resorting to explicit verbalization. For instance, Vosniadou and Brewer (1992) elicited drawings from 3rd and 5th 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.

As for problem solving, studies have shown that eliciting drawing production could provide useful insights with regards to the solvers' reasoning (Barrios & Martínez, 2014; Csíkos, Sztányi, & Kelemen, 2012; Edens & Potter, 2007, 2008). For instance, Edens and Potter (2008) instructed 4th and 5th graders to solve an arithmetic word problem and to make a drawing to help them find the solution. Using a custom scale, they graded to what extent the students' drawings were schematic or pictorial. They showed that the construction of schematic drawing was positively correlated with solving performance, and that most students (79%) rendered schematic representations. Similarly, drawings can be used to investigate solvers' mental representations while solving arithmetic word problems.

Another path to study solvers' representation of arithmetic word problems comes from prior work on strategy selection. Problems admitting multiple solving strategies are of particular interest, since the selection of one algorithm over another is informative about the representation constructed by the solvers (De Corte, Verschaffel, & De Win, 1985). For instance, Thevenot and Oakhill (2005, 2006) worked on a multiple-step problem solving task in which the cognitive load was manipulated through values size (large or small). They showed that depending on the size of the values, participants used different solving algorithms. The issue of the semantic determinants of problem representations can be tackled using such a

paradigm in which different solving strategies are available, and the solver's ability to pick and use one tells us about the abstracted interpreted structures (Gamo et al., 2010). For example, Coquin-Viennot and Moreau (2003) showed that the presence of a grouping element in a problem statement (such as flowers presented within a bouquet instead of separately) could incite participants to use a factorizing rather than a development algorithm.

Development of cardinality and ordinality

A key issue at the heart of our study regards the distinction between the perception of cardinality and ordinality in numerical situations. When children first learn how to count, they need to master a series of principles pertaining to the meaning of numbers (Gelman & Gallistel, 1986). Among those, it has notably been reported that children must learn the cardinal principle, that is, they have to understand that when they enumerate a series of entities in a collection, the count name they give to the last entity corresponds to the total number of entities in the collection. Another principle that children learn to master is the stable order principle, which indicates that the numbers used to enumerate items in a list need to be said in a specific, stable order. No matter what is being counted, the order in which the count names are stated remains the same. This principle refers to the ordinal property of numbers; their numerical position in an ordered sequence. This seminal work of Gelman and Gallistel (1986) highlighted the fact that humans need to learn that numbers can either refer to the general concept of quantity by stating the total number of elements within a set, or to the order in which the elements of an ordered list are sequenced.

Building upon this work, children's mastery of the cardinal principle has been investigated in several experimental works suggesting that children learn the cardinal meaning of numbers over the first few years of life (e.g. Bermejo, 1996; Condry & Spelke, 2008; Le Corre & Carey, 2007; Sarnecka & Lee, 2009; Wynn, 1992). 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 (e.g. "first", "second", "third") around the 4th and 5th year (Fischer & Beckey, 1990; Miller, Major, Shu, & Zhang, 2000; Miller, Marcovitch, Boseovski, & Lewkowicz, 2015). 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 (Colomé & Noël, 2012; Meyer, Barbiers, & Weerman, 2016; Wasner, Moeller, Fischer, & Nuerk, 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 learning how to count.

However, the influence played by the distinction between cardinality and ordinality in more advanced mathematical reasoning, involving other processes than simple enumeration, has seldom been studied. A first step in this direction was made by Gamo et al. (2010), who created arithmetic word problems made to emphasize either the cardinal aspect of numbers or their ordinal aspect.

Cardinality and ordinality in arithmetic problem solving

By using either number-of-elements problems in which distinct, unordered quantities are counted, or duration problems and height problems in which the counted quantities can be represented along a numerical axis such as a timeline, Gamo et al. (2010) showed that 5th graders' interpretation of the problems varied, thus influencing the strategies they used to solve the problems. Indeed, in their experiments, participants could construct different representations of isomorphic problems. An axis-based representation of the situation allowed them to perceive the relevance of the shortest algorithm to solve the problems, whereas a set-based representation of the situation favored the use of a longer, 3-step solving algorithm. Consider, for instance, the following problem:

There are 5 people in the Richard family. When the Richards go on holidays with the Roberts, they make a total of 14 people at the hotel. The Roberts are joined on holiday by the Dumas family. In the Dumas family, there are 3 people less than in the Richard family. The Roberts are going on holidays with the Dumas. How many will they be at the hotel?

Because this problem involves distinct entities (family members) that can be counted in any order, it is believed that most participants encode this problem as a cardinal, set-based representation (see Fig. 1). As a result, participants tend to consider that the only way to calculate the number of people at the hotel is to calculate the number of people in the Roberts family and add it to the number of people in the Dumas family. That is, most participants use a 3-step algorithm to solve this problem: $14 - 5 = 9$; $5 - 3 = 2$; $9 + 2 = 11$.

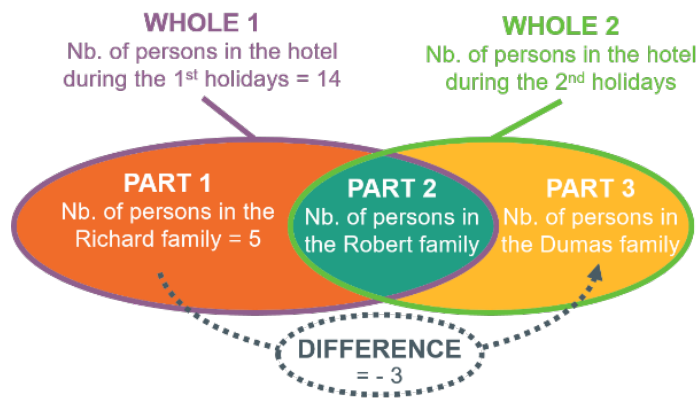


Fig. 1 Cardinal representation of problem 1. This representation fosters the calculation of the intersection (part 2) between whole 1 and whole 2, thus favoring the 3-step algorithm.

On the other hand, consider the following duration problem:

Antoine took painting classes for 5 years, and stopped at the age of 14. Jean started at the same age as Antoine, and went to classes 3 years less than him.

How old was Jean when he stopped attending painting classes?

This problem and the family problem share the same mathematical structure (see Fig. 2). However, because this second problem involves duration values instead of family counts, participants tend to represent the depicted situation along a timeline, with the numerical values being conceived of as states and transitions on an axis instead of as parts and wholes in a set-based representation (see Fig. 3). Constructing such a mental representation allows the solvers to see that there is a much shorter algorithm to solve the problem: instead of trying to calculate at which age Jean started taking painting classes, one may realize that since Jean and Antoine started attending the classes at the same age, and Jean stopped 3 years before Antoine, then Jean was 3 years younger than Antoine when he stopped attending the classes: $14 - 3 = 11$.

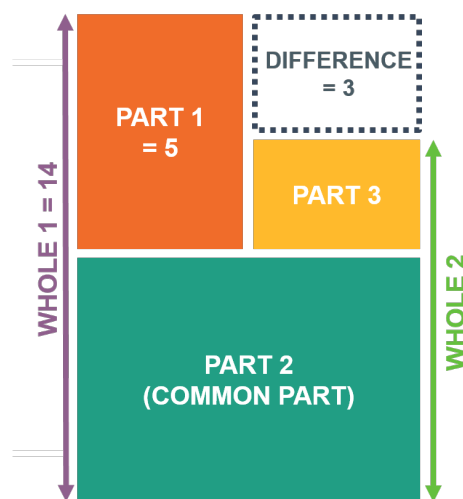


Fig. 2 Mathematical structure of the cardinal and ordinal problems.

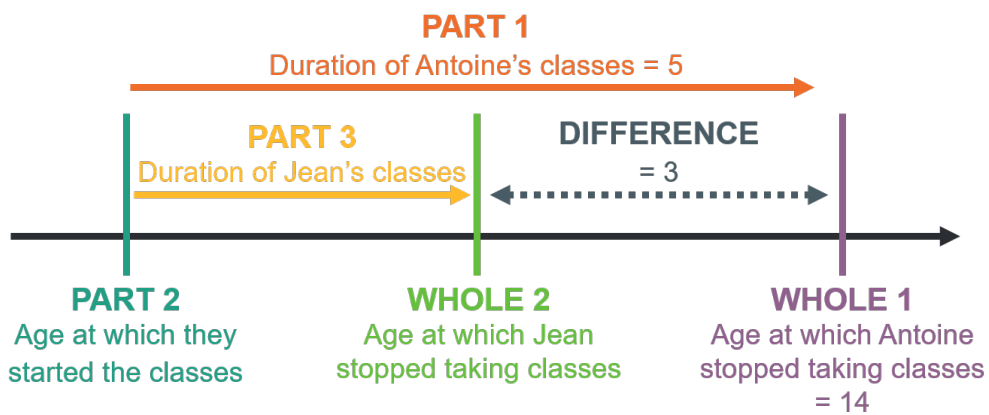


Fig. 3 Ordinal representation of the duration problem. This representation puts forward the fact that the difference between Whole 1 and Whole 2 is equal to the difference between Part 1 and Part 3. The shorter 1-step algorithm thus becomes easier to find.

Previous results on such problems have shown that participants have an easier time using the 1-step algorithm on ordinal problems than on cardinal problems (Gamo et al., 2010), to the point that even math experts may fail to recognize the validity of the 1-step algorithm on cardinal problems (Gros et al., 2019).

Present study

Our study built upon the works of Gamo et al. (2010) and Gros et al. (2019) in order to highlight the role played by the distinction between cardinal and ordinal quantities on children and adults' problem representation. We aimed at providing converging measures of the impact of this distinction on participants' ability to represent and solve the problems, and to provide the first empirical test of these effects on children and adults simultaneously. Additionally, by comparing children's and adults' performance on the same problems, we aimed at gathering insights into the developmental trajectory of these effects. Two experiments were designed to this end. First, we asked 5th graders and adults to consider a series of 12 problems. We asked participants to make a drawing of the problems and to try to solve them using as few operations as possible. The rationale was that cardinal problems would elicit a cardinal representation that would result in drawings with cardinal features and to a preferential use of the 3-step algorithm. On the other hand, we expected that ordinal problems would lead to the production of drawings with ordinal features and to the use of the 1-step algorithm instead. We used the same materials with children and adults to investigate how cardinal and ordinal representations develop over the years.

Second, we looked at the influence of the cardinal versus ordinal distinction on a younger population, by asking 2nd graders to solve a simplified version of the

problems. Two groups of students were recruited, for which different levels of arithmetic proficiency were expected. One group was composed of 8 classes of 2nd graders following a *business as usual* arithmetic program. The other group was composed of 8 classes of 2nd graders who had been receiving a special training in arithmetic (the *ACE program*, see Fischer, Sander, Sensevy, Vilette, & Richard, 2019), meant to improve their arithmetic reasoning skills. The arithmetic proficiency of this second group was hypothesized to be higher than that of the *business as usual* group, since the *ACE program* in charge of their arithmetic training had been shown to improve students' performance in a range of arithmetic-related activities, including arithmetic word problem solving (Fischer et al., 2019; Fischer, Vilette, Joffredo-Lebrun, Morellato, Le Normand, Scheibling-Sève, & Richard, 2019; Vilette, Fischer, Sander, Sensevy, Quilio, & Richard, 2017). Thus, by studying these two groups simultaneously, we hoped to gather evidence regarding the influence of the cardinal-ordinal distinction at different levels of arithmetic proficiency. We also tested the simplified problems on a population of adults to assess whether Experiment 1's findings could be replicated on easier problems, within reach of some 2nd graders. We predicted that even on simplified problems, the distinction between cardinal and ordinal quantities would lead to differences in algorithm use, from 2nd grade to adulthood.

Experiment 1

Methods

Participants. We recruited participants from two populations: a group of 59 children in 5th grade, recruited among several schools from the Paris region (27 females, $M = 11.00$ years, $SD = 0.36$), and a group of 52 adults from the Paris region (36 females, $M = 26.86$ 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: 6 using ordinal quantities (“duration”, “height”, and “number of floor” problems, see Table 1) and 6 using cardinal quantities (“collection”, “price”, and “weight” problems, see Table 2).

Table 8 Cardinal problems used in Experiment 1. The numerical values respected the following rule: $z < 4 < x < y < 15$.

Quantity used	Problem statement
---------------	-------------------

Weight	<p>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?</p>
Weight	<p>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?</p>
Price	<p>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 omelet with mushrooms, and an apple pie. The apple pie costs z euros less than chocolate cake. How much does the second meal cost?</p>
Price	<p>In the stationery shop, Antoine 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 Antoine, and an eraser. The eraser costs z euros less than the ruler. How much will Julie have to pay?</p>
Collection	<p>Paul has x red marbles. He also has blue marbles. In total, Paul has y marbles. Charlene has as many blue marbles as Paul, and some green marbles. She has z green marbles less than Tom has red marbles. How many marbles does Charlene have?</p>
Collection	<p>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?</p>

Table 9 Ordinal problems used in Experiment 1. The numerical values respected the following rule: $z < 4 < x < y < 15$.

Quantity used	Pb. ID	Problem statement
Duration	Pb. G	<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	Pb. H	<p>Sophie travels for x hours. Her trip started during the day. Sophie arrives at y. Fred leaves at the same time as Sophie. Fred's trip lasts for z hours less than Sophie's. What time is it when Fred arrives?</p>
Height	Pb. I	<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	Pb. J	<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>
Floors	Pb. K	<p>Naomi takes the elevator and goes up x floors. She left from the floor where her grandparents live. She arrives to the yth floor. Her brother Derek also takes the elevator from their grandparents' floor. He goes up z floors less than Naomi. What floor does Derek arrive to?</p>
Floors	Pb. L	<p>Karen takes the elevator and goes up x floors. She left from the floor where the gym is. She arrives to the yth floor. Yohan also takes the elevator from the floor where the gym is. He goes up z floors less than Karen. What floor does Yohan arrive to?</p>

Participants were given a 13-page booklet with instructions printed on the first page. The instructions stated that:

You will find an arithmetic problem on each page of this booklet. Your task is to solve the problems using as few operations as possible. You can use the ‘draft’ area, but please copy in the ‘response’ area all the operations that you used to come up with the solution. This is not a speed test: take your time to read and understand each of these problems. Remember that the goal is to solve the problems using as few operations as possible. For every problem, we ask you to write down every operation(s) that you used to come up with the solution, even the simplest one that you can calculate mentally. For instance, the computation $15 - 6 - 2 = 7$, should not be written as a unique operation, but broken down as $15 - 6 = 9$ and $9 - 2 = 7$, which then counts for two operations. When you have found the solution to a problem, please make a drawing that could help someone understand the problem and solve it. The goal is to give a visual representation displaying the different pieces of information needed to solve the problem. *Translated from French.*

The materials, instructions included, were identical for the adult and children participants. Each page in the booklet was divided into four parts: the problem statement, the “drawing” area, the “draft” area and the “response” area. Problem order was randomized across booklets.

Scoring. Participants’ drawings were analyzed using a custom scale, designed to evaluate to what extent they featured ordinal versus cardinal characteristics. The scale included 8 items: 4 cardinal criteria and 4 ordinal criteria involving the most relevant features of what would presumably be a prototypically ordinal (resp. cardinal) representation (see Fig. 4).

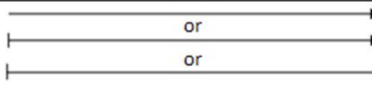
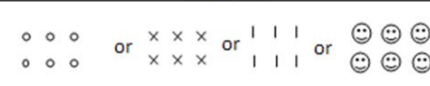
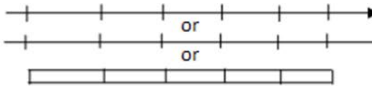

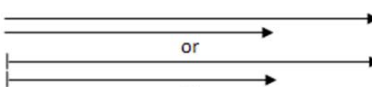
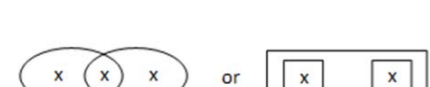


A. Ordinal criteria		B. Cardinal criteria	
Criteria	Examples	Criteria	Examples
Are there axes?		Are there groups of identical elements?	
Are there graduations?		Are there sets containing more than one element?	
Are there axes presented side by side?		Are there embedded sets?	
Are there intervals?		Are there correspondances between one element and one value?	

Fig. 4 Drawing scale used by the independent raters to grade the problems.

Double-blind scoring was performed by two independent raters who were unaware of the hypotheses being tested. The drawing scale was introduced to them and they were given the opportunity to ask questions about its different criteria and examples. Once they were confident they understood the meaning of each of the 8 items, they were asked to rate the entirety of the drawings produced by the participants in both groups. After initial scoring, the agreement between the two raters was 91.01%. Cohen's Kappa coefficient for inter-rater reliability was calculated to determine consistency among raters. The result ($\kappa = .726$, $SE = 0.012$) expressed substantial agreement between raters, according to Landis and Koch's typology (1977). After discussion, the raters reached 100% agreement on the drawings. Based on the raters' assessment of which criteria were met by each drawing, we calculated two scores: a cardinal drawing score, and an ordinal drawing score. The ordinal score (from 0 to 4) indicated how many of the 4 ordinal criteria were met by the drawings (see Fig. 4, column A). The cardinal score (from 0 to 4) indicated how many of the 4 cardinal criteria were met by the drawings (see Fig. 4, column B).

Participants' solving algorithms were deduced 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 *1-step algorithm* or as *3-step algorithm*. 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 labelled “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 algorithms (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. 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). Since every problem could be solved by both algorithms, solving algorithm analysis made it possible to identify which strategy was dominant in each context, and therefore to understand how the semantics attached to the problems related to participants’ use of solving algorithms.

Results

We analyzed the drawings produced by each participant. Fig. 5 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 type of problem, we compared participants’ mean cardinal and ordinal scores. As predicted, adults’ drawings on cardinal problems had a cardinal score ($M = 1.46$, $SD = 0.74$) higher than their ordinal score ($M = 0.04$, $SD = 0.10$); $t(51) = 13.87$, $p < .001$, $d = 2.69$, paired t-test. Similarly, their drawings of ordinal problems were rated higher on the ordinal scale ($M = 1.49$, $SD = 0.93$) than on the cardinal scale ($M = 0.54$, $SD = 0.36$); $t(51) = 7.42$, $p < .001$, $d = 1.34$, paired t-test. Thus, adult participants’ drawings met significantly more cardinal criteria when they drew cardinal problems, and more ordinal criteria when they drew ordinal problems.

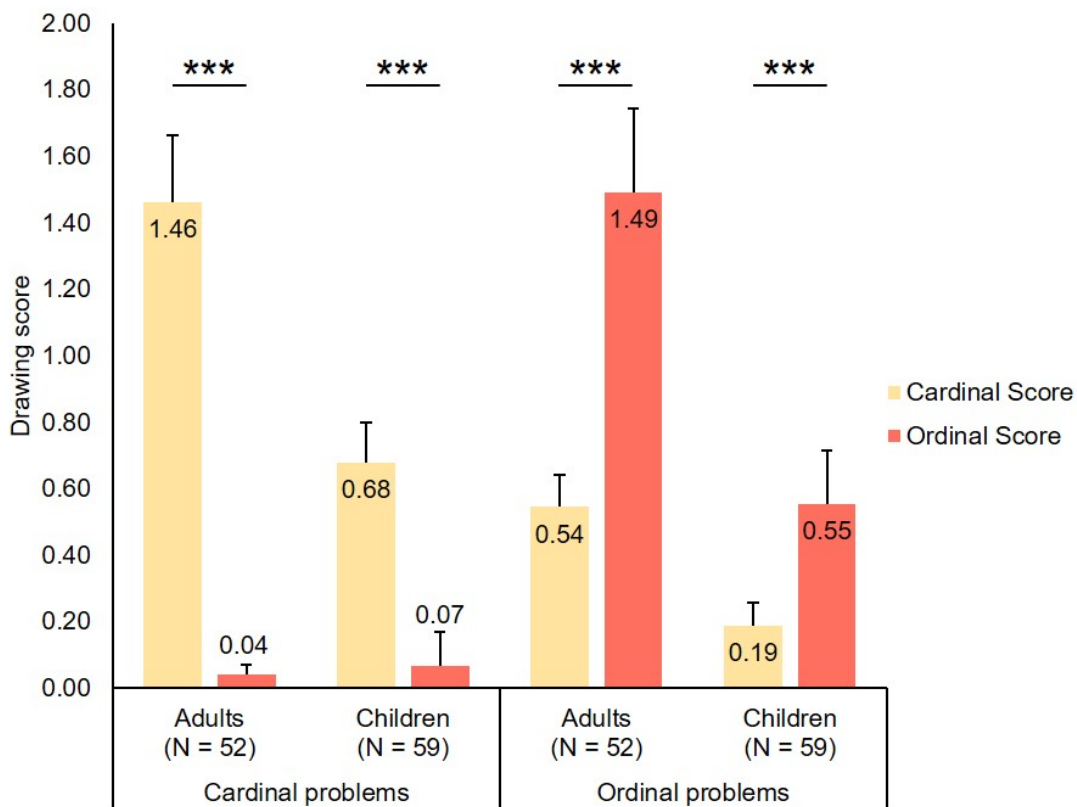


Fig. 5 Children’s and adults’ mean cardinal and ordinal scores depending on the type of quantity used in the problems. Vertical bars denote 0.95 confidence intervals. *** $p < .001$, paired t-test.

Regarding children, as predicted, their drawings on cardinal problems also scored higher on the cardinal scale ($M = 0.68$, $SD = 0.47$) than on the ordinal scale ($M = 0.07$, $SD = 0.39$); $t(58) = 8.27$, $p < .001$, $d = 1.41$, paired t-test. Similarly, the drawings they produced on ordinal problems had a higher ordinal score ($M = 0.55$, $SD = 0.63$) than cardinal score ($M = 0.19$, $SD = 0.27$); $t(58) = 5.82$, $p < .001$, $d = 0.75$, paired t-test. Thus, young participants were also more likely to use ordinal features (axes, graduations, intervals) than cardinal features (sets, groups of unordered elements, etc.) on ordinal problems, and the conversely on cardinal problems. In sum, in both populations, the presence of ordinal (resp. cardinal) quantities seems to result in representations featuring a higher number of ordinal (resp. cardinal) features, in both children and adults.

Second, we made the prediction that problems with ordinal quantities would facilitate the use of the 1-step algorithm compared to problems with cardinal quantities. In both groups, we evaluated whether participants did use the 1-step algorithm more often on problems involving ordinal quantities than on problems involving cardinal quantities. Fig. 6 details the participants’ use of each algorithm

depending on the type of quantity featured in the problems. A paired t-test analysis revealed that the mean rate of use of the 1-step algorithm was higher on ordinal ($M = 0.39$, $SD = 0.31$) than on cardinal ($M = 0.08$, $SD = 0.17$) problems; $t(58) = 8.36$, $p < .001$. On the other hand, participants used the 3-step algorithm more frequently on cardinal problems ($M = 0.33$, $SD = 0.35$) than on ordinal problems ($M = 0.12$, $SD = 0.19$); $t(58) = 5.64$, $p < .001$.

The same analyses were performed for the adults and showed that the mean rate of use of the 1-step algorithm was also higher on ordinal ($M = 0.46$, $SD = 0.33$) than on cardinal ($M = 0.25$, $SD = 0.35$) problems ($t(51) = 4.99$, $p < .001$, $d = 0.69$). Similarly, the rate of use of the 3-step algorithm was higher on cardinal problems ($M = 0.61$, $SD = 0.36$) than on ordinal problems ($M = 0.41$, $SD = 0.32$); $t(51) = 4.63$, $p < .001$. This confirmed that the choice of a solving algorithm is influenced by the cardinal versus ordinal nature of the problem's quantities, and that this effect is robust among both populations. Additionally, the 1-step algorithm was significantly less used by children than by adults on cardinal problems ($t(109) = 3.48$, $p < .001$, unpaired t-test) but not on ordinal problems ($t(109) = 1.10$, $p = .27$, unpaired t-test).

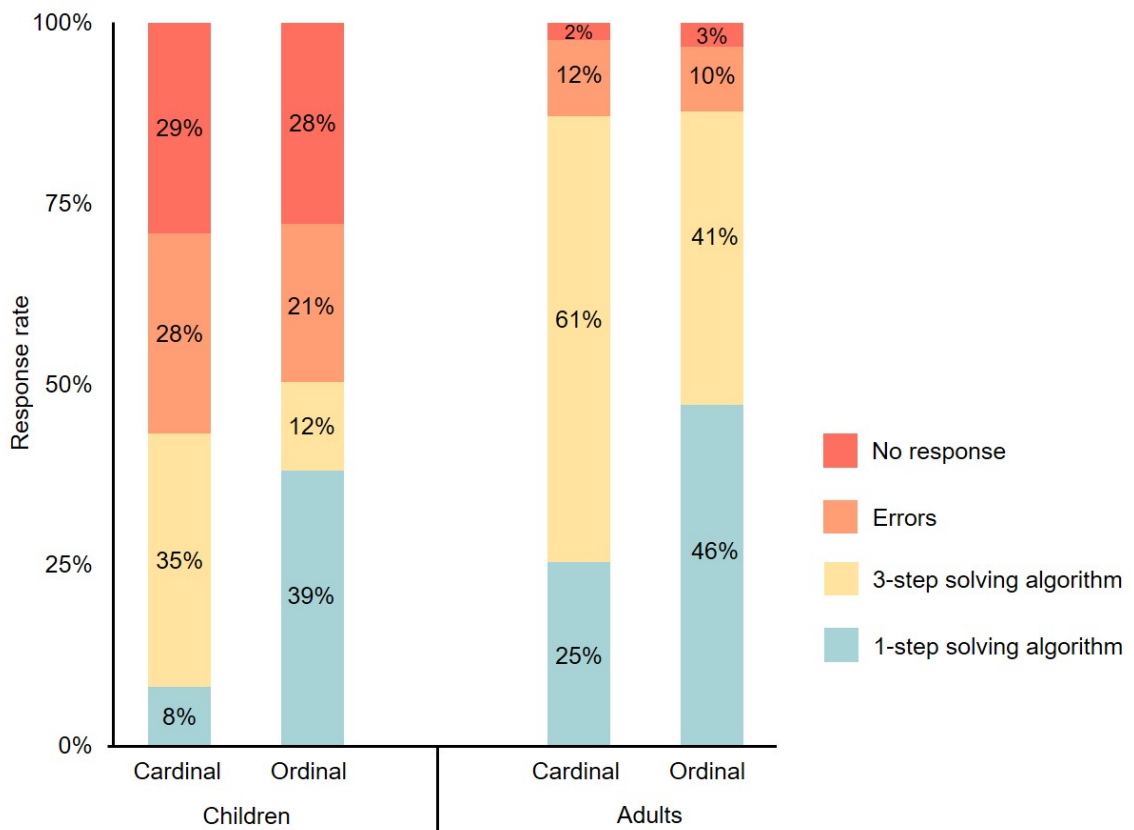


Fig. 6 Children's and adults' mean rate of use of the two solving algorithms depending on the quantities used in the problems.

Third, we looked at how the ordinality of the drawings correlated with participants' use of the 1-step algorithm, 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 the success in using the 1-step algorithm 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 to account for the design's repeated measures. Among children, there was a significant effect of the ordinal drawing score on participants' use of the 1-step algorithm, even when accounting for the effect of the cardinal versus ordinal nature of the problems ($z = 2.44, p < .05$). In other words, a higher ratio of ordinal features in the drawings predicted higher chances to use the 1-step algorithm among children. Among adults, however, contrarily to our hypothesis, there was no significant effect of the ordinal drawing score on the rate of use of the 1-step algorithm ($z = 0.99, p > .1$).

Discussion

By resorting to a drawing task coupled with a solving task, we were able to get new insights into the participants' representation of arithmetic word problems. The drawing analysis allowed us to get a complementary look at participants' conceptualization of cardinal and ordinal problems without resorting to explicit verbalization. Results supported the idea that cardinal problems foster cardinal representations involving sets and unordered collections, whereas ordinal problems favor the construction of ordinal 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. Additionally, Gamo et al.'s (2010) finding regarding algorithm choice was replicated in this experiment, since participants in both groups tended to use the 1-step algorithm on ordinal problems whereas they preferentially used the 3-step algorithm on cardinal problems, regardless of the instructions asking them to solve the problems using as few operations as possible. Finally, the ordinality of children's drawings was predictive of their tendency to use the 1-step algorithm to solve the problems – regardless of whether the problems were cardinal or ordinal to begin with – which seems to indicate that the features we were attentive to in the drawings were

relevant to explain participants' reasoning process. In other words, a drawing featuring axes instead of sets was more likely to indicate that children would find the 1-step algorithm. This effect, however, was not replicated among adults. Their ordinal drawing scores on cardinal problems were slightly higher, on average, for problems successfully solved with the 1-step algorithm ($M = 1.56$, $SD = 1.32$) than for problems for which they did not find the 1-step solution ($M = 1.44$, $SD = 1.16$), but this difference was not significant. This might be due to adults scarcely using ordinal features on their drawings of cardinal problems (0.04 ordinal criteria on average), despite them finding the 1-step algorithm for 25% of those problems. It may be possible that due to their extensive habit of representing cardinal quantities as cardinal representations, adults reverted to their initial encoding of the problems when asked to make a drawing of the problems. Further work is needed on this question to elucidate the reason behind this null effect.

Taken together, the results on both tasks suggest that the non-mathematical knowledge about the elements featured in the problems has a pervasive influence on children and adults alike, leading them to abstract either a cardinal or an ordinal representation of the situations and thus shaping their solving strategies. In order to gather evidence regarding the pervasiveness of these encoding effects at all ages, we designed a second experiment looking at an even younger population.

Experiment 2

Since a central aspect of learning how to count in the first years of life resides in understanding the cardinal and ordinal properties of numbers, we decided to investigate the influence of the cardinal-ordinal distinction on a younger population. We hoped to shed some light on the way in which this distinction manifests itself in word problem solving before the 5th grade. We recruited 2nd graders and asked them to solve a series of arithmetic word problems. In order to adapt our materials to the lower arithmetic word problem solving proficiency of 2nd graders, we created new, simplified versions of the problems used in Experiment 1.

As previously mentioned, this experiment was conducted as part of the assessment of the experimental arithmetic teaching program ACE ArithmEcole, aimed at improving 2nd graders' understanding of mathematics. A total of eight classes participated in the program and were then asked to complete the task presented in this experiment. The data collection occurred at the beginning of May, which means

that the students had been enrolled in the ACE program for 8 full months before taking our test. We also recruited 8 classes that did not participate in the program to compare their results with the performance of the program's students. While the evaluation of the ACE program is not the purpose of this article, we analyzed data from ACE students separately from data collected with non-ACE students, since an increase in mathematical proficiency was expected in the ACE group. Because we created new simplified problems for this experiment, we also recruited a group of adult participants to compare how they fared on the new, easier problems.

Methods

Participants. A total of 193 students from the ACE program were recruited for this experiment, coming from 8 different 2nd grade classes ($N = 193$, 92 girls, $M_{\text{age}} = 7.44$ years, $SD = 0.33$ years). We also recruited 191 students from 8 2nd grade classes that did not participate in the program but that matched those of the ACE group on socio-economic criteria ($N = 191$, 103 girls, $M_{\text{age}} = 7.45$ years, $SD = 0.34$ years). Finally, the adult group comprised 70 participants recruited online, through social networks and mailing lists. Three adults interrupted the experiment before completion and were excluded from the subsequent analyses ($N = 67$, 41 women, $M_{\text{age}} = 29.31$ years, $SD = 13.02$ years).

Materials and procedure. Considering the early age of our participants in this second experiment, we modified the problems used in Experiment 1 to make them easier to solve and understand, while still preserving their abstract mathematical structure. We created four simplified problems (see Table 3). Instead of the 6-sentence long problems used in Experiment 1, these were only 3-sentence long. Additionally, the value of Part 3 was higher than the value of Part 1, so that the 1-step algorithm consisted in a single addition (e.g. $10 + 2 = 12$) instead of the subtraction in Experiment 1. Numerical values were randomized between problems and between participants. The value of Whole 1 was comprised between 9 and 12, the value of Part 1 between 3 and 5, and the value of the Difference was always 2.

Regarding the two student groups, the experiment took place in their classrooms, during normal class hours. Students were presented with a booklet on which the four problems were printed on separate pages. Instructions were given orally to the second graders:

You will find a math problem on each page of this booklet. Your task is to solve the problems using as few operations as possible. Write down all the

operations that you used to come up with the solution. This is not a speed test: take your time to read and understand each of these problems. Remember that the goal is to solve the problems using as few operations as possible. For every problem, we ask you to write down every operation that you used to come up with the solution, even the simplest one that you can calculate mentally. *Translated from French.*

Participants were given 45 minutes to complete the task. Each problem was read out loud twice by the experimenter, and the students were then instructed to read it by themselves and to try to solve it.

Table 10 Simplified cardinal and ordinal problems used in Experiment 2 (translated from French).

A. Simplified ordinal problems	B. Simplified cardinal problems
<p>Mark's train arrived at 10: he had travelled for 3 hours. Judith took the train for 2 more hours than Mark, after leaving at the same time as he did. When does Judith arrive?</p>	<p>Sarah has 10 pets: 3 cats, and dogs. Bob has 2 cats more than Sarah, and as many dogs as her. How many pets does Bob have?</p>
<p>Matteo took piano lessons until he was 10: he attended the class for 3 years. Lisa took piano lessons for 2 more years than Matteo, and she started at the same age as he did. How old was Lisa when she stopped attending the piano lessons?</p>	<p>Paul has 10 marbles: 3 blue marbles, and some red marbles. Zoey had 2 blue marbles more than Paul, and as many red marbles as he does. How many marbles does Zoey have?</p>

Regarding the adult group, the experiment took place online, on the Qualtrics platform for online experiments. Participants were sent an anonymous link to complete the experiment online. The first page displayed the same instructions that were given to the second graders. Participants were told to solve the problems using as few operations as possible, and to write down every calculation they made in the

process. Participants navigated the experiment themselves, using the “next” button to go from one problem to another.

Results

We were interested to see if the results of Experiment 1 could be replicated with younger participants, and with simpler problems. We scored the algorithms used by the participants using the same rules as in Experiment 1, and we investigated how often pWe aimed at replicating Experiment 1 with younger participants and with simpler problems. To this end, we scored the algorithms used by the participants using the same rules as in Experiment 1, and we investigated how often participants used the 1-step algorithm on cardinal and on ordinal problems (see Fig. 7). We had predicted that the difference between cardinal and ordinal problems would result in a difference in algorithm use in the three experimental groups. In accordance with our prediction, the 2nd graders who did not attend the ACE program used the 1-step algorithm significantly more often on ordinal problems (in 22.77% of trials) than on cardinal problems (in 13.35% of trials); $t(190) = 3.57, p < .001, d = 0.26$, paired t-test. Interestingly, the 3-step algorithm was extremely rarely used by this population. In fact, no participant of the non-ACE group ever used the 3-step algorithm on ordinal problems, and they used it in only 4.45% of the trials on cardinal problems; $t(190) = 3.84, p < .001$, one-sample t-test against zero. The relatively low rate of use of both the 1-step algorithm and the 3-step algorithm is in part attributable to the high number of errors made by the participants of the non-ACE group (74.21% wrong answers). As expected, the 2nd graders who attended the ACE program achieved higher performances on average than the 2nd graders who did not attend the program: they used the 1-step algorithm in 28.76% of the trials, whereas the non-ACE students only used the 1-step algorithm on 18.06% of the problems, $t(382) = 3.88, p < .001$, unpaired t-test. However, the 2nd graders who benefitted from the ACE program still used the 1-step algorithm more frequently on ordinal problems (36.27% of the cases) than on cardinal problems (21.24%), as predicted; $t(192) = 5.23, p < .001, d = 0.39$, paired t-test. In other words, the effect of the cardinal-ordinal distinction remained significant despite the higher proficiency of the ACE 2nd graders. Similarly, they used the 3-step algorithm more often on cardinal problems (8.55%) than on ordinal problems (only 1 occurrence, corresponding to 0.26% of the trials). A one-sample t-test confirmed that the rate of use of the 3-step algorithm on cardinal problems was significantly higher than zero; $t(192) = 4.35, p < .001$, one-sample t-test.

Regarding the adults, as could be expected, their overall performances were better than those of the students in both groups, since they managed to find the 1-step algorithm in 68.28% of the trials, which was significantly higher than ACE students' performance; $t(258) = 9.00, p < .001$, unpaired t-test. Contrarily to the two student groups, adults' use of the 1-step algorithm did not significantly differ between cardinal problems (66.42%) and ordinal problems (70.15%); $t(66) = 0.63, p = .53$; paired t-test. However, they used the 3-step algorithm significantly more often on cardinal problems (33.58%) than on ordinal problems (16.42%); $t(66) = 2.83, p < .01$, paired t-test. Thus adults' strategy use differed between cardinal and ordinal problems, but only with regards to the 3-step strategy.

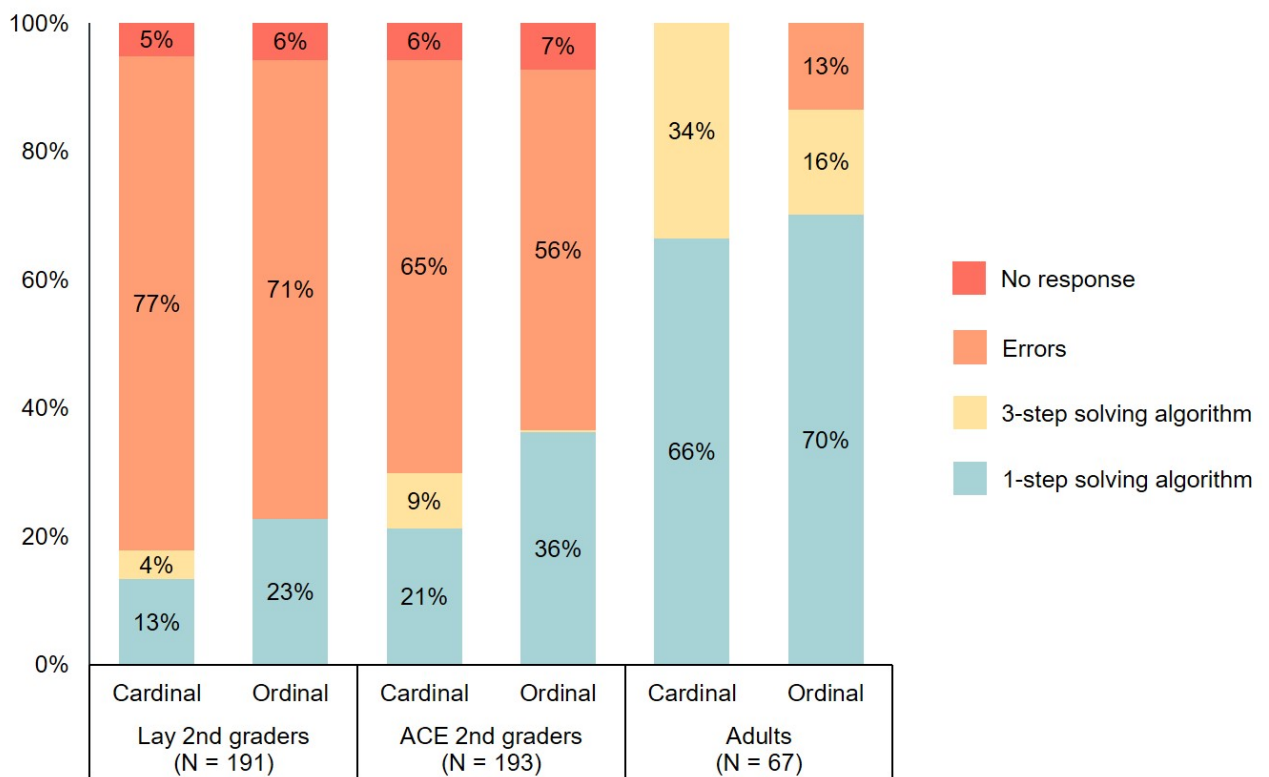


Fig. 7 Algorithm use depending on the cardinal versus ordinal nature of the problems.

Discussion

This second experiment confirmed that the use of cardinal versus ordinal quantities within a problem statement had a robust influence on solvers' use of solving strategies, fostering the 1-step algorithm on ordinal problems and the 3-step algorithm on cardinal problems. Even among a population of 2nd graders, and despite their high rate of failure on the problems, they still managed to use the 1-step algorithm more often on duration problems than on collection problems. The student group who followed a special ACE training was overall better than the group who did not, but

the difference between cardinal and ordinal problems remained present. In other words, participants' ability to solve the problems improved, but they were still influenced by the cardinal or ordinal context in which the problems were embedded.

Interestingly, the 3-step algorithm was seldom used by the students, whereas in Experiment 1 the fifth graders used it in most cases to solve cardinal problems. This difference might be attributable to the changes introduced in the problem statements: in an attempt to make the problems easier to solve with the 1-step algorithm, we changed the sign of the difference (e.g. "Bob has 2 *more* cats than Sarah"), whereas the problems in Experiment 1 systematically introduced a negative difference. Because of this change, the 1-step algorithm consisted in a single addition: (e.g. " $10 + 2 = 12$ "), whereas the 3-step algorithm still involved one subtraction (" $10 - 3 = 7$; $3 + 2 = 5$; $7 + 5 = 12$ "). Thus, since the notion of subtraction is usually introduced in 2nd grade in France, it might be that the 3-step algorithm was simply too difficult to implement at this age. Also, the low rate of use of the 3-step algorithm by the 2nd graders might in part be due to the 3 necessary steps to find the solution. Multiple-steps arithmetic problems are not a common occurrence at this age, and students who did not manage to find the 1-step algorithm may have simply lacked the resource to plan and compute a 3-step solution strategy.

Since the problems we used were not identical to those used in Experiment 1, we also included a group of adults who were presented the same task as the 2nd graders. As could be expected, adults' performance were significantly higher than that of 2nd graders on the task. However, contrary to Experiment 1, the difference in rate of use of the 1-step algorithm disappeared among adults, which may be explained by the fact that the simplified problems made it easier for adult participants to recode their initial problem representation to find the 1-step algorithm (they succeeded in using the 1-step algorithm in 66% of the cases, whereas they had only reached 25% in Experiment 1). Interestingly, participants still tended to use the 3-step algorithm more often on cardinal problems than on ordinal problems, suggesting that there was still an effect of the use of cardinal versus ordinal quantities in the problems. Thus, despite their indisputable expertise on math problems within reach of around a fifth of 2nd graders, the adults still fell prey, to a certain extent, to interferences between the problems' mathematical structure and the semantics conveyed by the situations depicted in the problem statements.

General discussion

Overall, the results of these two experiments show that the use of cardinal and ordinal quantities within problem statements has a decisive influence on participants' encoding of the problems. Students' apprehension of marbles, years, cats or trains led them to construct semantically dissimilar representations of the situations depicted: the distinction introduced between ordinal and cardinal problem statements was reflected in the constructed representations (as shown by the drawings the participants made) and resulted in them using different solving algorithms, even when specifically asked to use the shortest strategy they could find. Furthermore, the fact that those effects could be highlighted with high and low-proficiency 2nd graders, 5th graders, as well as adults indicates the pervasiveness 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 present.

In fact, despite recent evidence showing that children first learn about numbers' cardinal aspect before understanding their role as ordinal entities, it appeared in our experience that problems evoking an ordinal representations led to higher task performances (students were more likely to find the shortest solving algorithm) even among the youngest and less proficient population we tested. These results point in favor of an early development of the ability to construct ordinal representations of numerical situations. It may be that children's experience of the world lead them to perceive the order in specific quantities and to develop the ability to think in terms of sequences when the situations they encounter have salient ordinal properties.

Overall, the use of a double measure of the influence of the solvers' knowledge about the world allowed us to gather converging clues shedding light both into the abstracted representations and into the algorithms subsequently implemented. By focusing on the role of semantic properties on the initial encoding of a problem, we hope to gain a finer understanding of arithmetic problem solving as a whole, and to help describe the interactions between world semantics and mathematical semantics in arithmetic word problem solving. Understanding the determinants of problems' representations is a crucial step to identify the potential pitfalls and dead ends born from semantic incongruence, as well as to help develop analogical transfer in a

scholastic setting, by promoting semantic recoding among the learners (Gamo et al., 2010; Gros, Thibaut, & Sander, 2015).

References


- 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., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology*, *35*, 99-134
- Bassok, M. (2001). Semantic alignments in mathematical word problems. In Gentner, Holyoak and Kokinov (eds.) *The analogical mind: Perspectives from cognitive science*, (pp.401-433). Cambridge, Ma: MIT Press
- Bassok, M., Pedigo, S. F., & Oskarsson, A. T. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 343.
- Bassok, M., Wu, L., & Olseth, L. K. (1995). Judging a book by its cover: interpretative effects of content on problem-solving transfer. *Memory and Cognition*, *23*, 354 e 367.
- 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.
- Colomé, À., & 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.
- Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, *137*(1), 22-38.
- 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.
- Csíkós, 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.

- 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.
- 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.
- 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.
- Fischer, J. P., Vilette, B., Joffredo-Lebrun, S., Morellato, M., Le Normand, C., Scheibling-Sève, C., & Richard, J. F. (2019). Should we continue to teach standard written algorithms for the arithmetical operations? The example of subtraction. *Educational Studies in Mathematics*, 101(1), 105-121.
- Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, 34(8), 1430-1451.
- 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.
- Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number*. Cambridge, MA, US: Harvard University Press.
- 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*, online first. doi: 10.3758/s13423-019-01628-3.
- Gros, H., Thibaut, J.-P., & Sander, E. (2015). Robustness of semantic encoding effects in a transfer task for multiple-strategy arithmetic word problems. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 818-823). Austin, TX: Cognitive Science Society.

- Guthormsen, A. M., Fisher, K. J., Bassok, M., Osterhout, L., DeWolf, M., & Holyoak, K. J. (2016). Conceptual integration of arithmetic operations with real-world knowledge: Evidence from event-related potentials. *Cognitive science*, 40(3), 723-757.
- Ischebeck, A., Schocke, M., & Delazer, M. (2009). The processing and representation of fractions within the brain: An fMRI investigation. *NeuroImage*, 47(1), 403-413.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological review*, 92(1), 109.
- 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*, 159-174.
- 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.
- Meyer, C., Barbiers, S., & Weerman, F. (2016). Order and Ordinality: The Acquisition of Cardinals and Ordinals in Dutch. In *Boston University Conference on Language Development (BUCLD) 40*. Cascadilla.
- 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), 129-140.
- 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.
- Pinnegar, S., Mangelson, J., Reed, M., & Groves, S. (2011). Exploring preservice teachers' metaphor plotlines. *Teaching and Teacher Education*, 27(3), 639-647.
- Riley, M. S., Greeno, J. G., & Heller, J. 1. (1983). Development of children's problem-solving ability in arithmetic. *The development of mathematical thinking*, 153-196.
- Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology*, 103(3), 325-337.
- 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.
- Vilette, B., Fischer, J. P., Sander, E., Sensevy, G., Quilio, S., & Richard, J. F. (2017). Peut-on améliorer l'enseignement et l'apprentissage de l'arithmétique au CP ? Le dispositif ACE. *Revue française de pédagogie. Recherches en éducation*, (201), 105-120.
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive psychology*, 24(4), 535-585.
- 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.
- Wundt, W. (1907). Über Ausfrageexperimente und über die Methoden zur Psychologie des Denkens. *Psychologische Studien*, 3, 301-360.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358(6389), 749-750.

CHAPTER 7 – ROBUSTNESS OF SEMANTIC
ENCODING EFFECTS IN A TRANSFER TASK
FOR MULTIPLE-STRATEGY ARITHMETIC
PROBLEMS



Presentation

Learning how to use a mathematical notion in situations where it is relevant is the aim of any form of mathematics education. However, in order to do so, one needs to be able to apply what they learnt in a specific situation to new situations more or less similar to the initial one. It is this idea of *transfer* that we explore in this last experimental chapter. The previous chapters have established that different encodings are constructed when attempting to solve cardinal and ordinal word problems. This raises the question of what it takes to be able to perceive the similarity between two problems evoking dissimilar encodings but sharing the same mathematical structure. Chapter 2 showed that eliciting comparisons between isomorphic problems was not enough to overcome the differences arising from semantic incongruence. In this brief chapter, we explicitly informed participants that cardinal and ordinal problems can be solved according to the same principle, and to assess whether students can use this information to transfer a solving algorithm from one problem to another.

The experiment reported in this chapter was conducted among 110 fifth and sixth graders. They were presented with 2 ordinal training problems, for which the solutions and algorithms were provided, followed by 8 test problems. The solution to the two training problems was given to the participants, and they were instructed to use the same strategy to solve the test problems. Some of the test problems had the exact same mathematical structure as the training problems, whereas other test problems had been slightly modified so that the solving algorithm provided on the training problems would need to be adjusted. We predict that the influence of the quantities used in the problem statements is so robust that transferring the solution to an ordinal test problem will be easier than transferring it to a cardinal test problem, regardless of the modifications introduced between the training and the test problems (change of operands, change of operator). Additionally, we predict that transfer errors will be mainly due to the superficial changes introduced between the training and the test problems: a change of operand between the two problems will result in an error in operand choice, etc. In other words, we predict that failing to solve the test problems will be associated with a superficial application of the training problem's solving algorithm. This experiment suggests that providing participants with a worked-out example is not enough to help them engage in semantic recoding.

This article has been published in the *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (2015). It is reproduced as is in the following chapter. Please note that since it was the first article written at the beginning of this PhD, the terminology used may differ from the one used in the rest of this thesis.

Abstract

The nature of the quantities involved in arithmetic problems promotes semantic encodings that affect the strategy chosen to solve them. Such encoding effects might prevent positive transfer to problems sharing the same formal mathematical structure. In this study with 5th and 6th graders, we investigated the conditions promoting positive and negative transfer in arithmetic problems that could be solved with two distinct strategies. We showed that basic training cannot overcome the initial impact of semantic encodings, and we provided evidence that a lack of semantic encoding of the training problems leads to transfer errors. This suggests the existence of ontological restrictions on the representation mechanisms involved in problem solving tasks.

Keywords

arithmetic problem solving · analogical transfer · semantic structures · semantic alignment
· strategy choice

Introduction

Semantic content in arithmetic problem solving. It is well established that the semantic content of arithmetic word problems influences their difficulty. For example, among one-step subtraction problems, when the question bears on the final result, change problems (e.g., “John had 8 marbles, he loses 5 marbles during recess. How many marbles does John have now?”) are easier than combine problems (e.g., “John and Tom have 8 marbles altogether, Tom has 5 marbles. How many marbles does John have?”) (Riley, Greeno, & Heller, 1983). In the case of conceptual rewording, when semantic cues relevant to the solution are provided, the construction of an appropriate mental representation is facilitated and the problem is easier to solve (Vicente, Orrantia, & Verschaffel, 2007). Success depends on a process of semantic alignment that aligns semantic relations evoked by the entities of the problem situation with mathematical relations (Bassok, Chase, & Martin, 1998).

Change in encoding and choice of strategy. A problem can be described in terms of semantic dimensions, such as static state versus dynamic process or discrete versus continuous states which, in turn, influence the representation of the problem and the solution strategies (De Corte et al., 1985; Bassok & Olseth, 1995). The encoding might influence not only the problem difficulty but also the solving strategy (Brissiaud & Sander, 2010). Furthermore, one specific encoding of the problem might be more efficient than another in terms of number of steps necessary to reach the solution. This is the case for distributive word problems (Coquin-Viennot & Moreau, 2003) or multiple-step arithmetic word problems (Thevenot & Oakhill, 2005). For example, in Coquin-Viennot and Moreau (2003), Grade 3 and 5 pupils were given problems that could be solved by a distributed strategy (e.g., $k \times a + k \times b$), or a factorized strategy [e.g., $k \times (a + b)$]; the presence of a word cueing the grouping of elements increased the frequency of the factorized strategy.

Gamo, Sander & Richard (2010) showed that depending on the type of quantities used in arithmetic problems, the emphasis put on one of the two following relationships will be different; (1) the complementation relation primes the computation of the difference between a whole and one of its component parts, or (2) a matching relation leading to the computation of the difference between homologous quantities. Compare, in this respect, the two following problems (a) “In the Richard family, there are 5 persons. When the Richards go on vacation with the

Roberts, they are 9 at the hotel. In the Dumas family, there are 3 fewer persons than in the Richard family. The Roberts go on vacation with the Dumas. How many will they be at the hotel?” and (b) “Antoine took painting courses at the art school for 8 years and stopped when he was 17 years old. Jean began at the same age as Antoine and took the course for two years less. At what age did Jean stop?” Both can be solved by the same two strategies. However, most participants solve the first one with a complementation strategy (i.e., $9 - 5 = 4$; $5 - 3 = 2$; $4 + 2 = 6$) and almost never use the matching strategy ($9 - 3 = 6$) whereas in the second type of problem a majority of participants use the matching strategy ($17 - 2 = 15$) and much less the complementation strategy ($17 - 8 = 9$; $8 - 2 = 6$; $9 + 6 = 15$) (Gamo et al., 2010). The fact that, in the first case, the question refers to the total number of persons and that a component family is missing seems to imply that this set should be calculated first. This is a case of cardinal encoding. By contrast, in the second example, the problem triggers an ordinal encoding which makes it more salient that the difference in terms of course duration is equivalent to the age difference between Antoine and Jean at course completion. Thus, the age at which Jean stopped can be obtained by subtracting their age difference. In order to use the matching strategy, a recoding is necessary to infer that the difference in the number of persons who do not participate in both trips, that is, the Richard and the Dumas families, reflects the difference between the number of persons within the two groups (Gamo et al., 2010). These problems are compatible with two semantic alignments: the semantic relations evoked by the entities of the problem situation might be aligned with two kinds of mathematical relations (complementation or matching relations), each one associated with a different solving strategy that leads to the correct solution (complementation or matching strategy). A cardinal encoding emphasizes the complementation relations and an ordinal encoding emphasizes the matching relations.

The semantic determinants of transfer. Transfer has been shown to be more effective when surface features –those that can be manipulated without modifying the solution or the solving procedures– from the source and the target problems are similar (e.g. Novick & Holyoak, 1991). Bassok and Olseth (1995) showed that surface features were not only interfering with the structural ones, but were inducing a semantic structure that could be more or less congruent with the mathematical one. Surface features appear to be instantiations of abstract semantic dimensions such as symmetry-asymmetry. Analogical transfer was shown to be influenced by these

dimensions. Permutation problems with symmetric sets of elements (e.g., doctors from Chicago and doctors from Minnesota are symmetric because they have equivalent semantic roles in the world) were not considered to be of the same type as permutation problems with asymmetric sets (e.g., prizes and students, in which, in the real world, prizes are given to students rather than students are given to prizes). As a consequence, being trained on one type of problem or on another influences the performance on the test problems (Bassok, Wu, & Olseth, 1995).

Goal of the present study. Most of the studies on transfer used problems in which there is only one successful strategy. Thus, failures to transfer were confounded with failures to solve the problem. However, these failures might correspond to quite different cases. Failures might result from a poor representation of the problem or from failures to match appropriately the source and the target despite the existence of an adequate representation of the problem. Thus, it is difficult to distinguish between representational aspects and strategic ones. By contrast, in the current study, we used arithmetic problems that could be correctly solved with the two distinct strategies presented above, the complementation strategy (3 steps) or the matching strategy (1 step). It is thus possible to dissociate positive transfer resulting from the strategy taught in the source problem from a successful resolution based on the other available strategy which also leads to a correct solution. The latter relies on another representation of the problem than the one that would lead to transfer of the strategy.

In the present study, participants know the algorithms and their mathematical meaning (i.e., they know how and what it means to add or to subtract, and what it means to look for the value of a part or a whole, or to compare quantities). The main point is to study whether they will be able to transfer a new solving strategy in various contexts.

By contrast, in most previous experiments, the problems, such as permutation problems (e.g. Ross, 1989; Bassok & Olseth, 1995), were quite complex. Thus, the origin of transfer failures remains unclear. Did participants understand the meaning of the algorithms they were provided with? It is possible that they “blindly” applied the algorithms from the source problem with very poor understanding of the underlying mathematical features. If they failed to understand the meaning of the algorithms, they might have mapped the training problem on the transfer items on the basis of perceived equivalence of roles (i.e., this entity in the training problem has the same role as that entity in a transfer problem, so I give them the same role in

the algorithm). The main source of failure would then be the inappropriate encoding of the training situation.

We hypothesized that the transfer of the matching strategy to novel problems sharing the same formal mathematical structure should be influenced by the type of representation induced by the problems. In that respect, we trained pupils on examples of the matching (1 step) strategy, and then asked them in a test phase to use it in several types of problems, which varied with respect to their similarity to these example problems. We decided to teach the matching strategy only. Indeed, there will be no reason to teach the complementation strategy, which is less efficient and which is spontaneously used for cardinal problems. For ordinal problems, they would spontaneously choose the matching strategy and they would be no point, in this case, to use the 3-step, less efficient complementation strategy.

Presentation of the problems. All of the problems had the same formal mathematical structure as the ones used in Gamo et al. (2010), presented in Fig. 1.

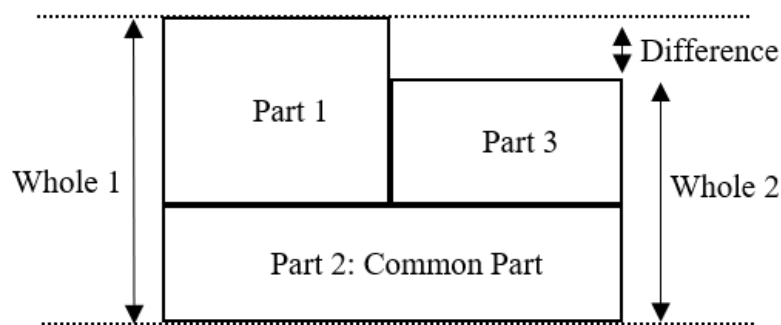


Fig. 1 Formal mathematical structure of the problems.

In the previous examples, the Roberts (respectively the age at which Jean and Antoine began their course) represent Part 2, the Richards (respectively the duration of the course for Antoine) represent Part 1, the Dumas (respectively the duration of the course for Jean) represent Part 3, the Roberts and the Richards (respectively the age at which Antoine stopped his courses) represent Whole 1, the Roberts and the Dumas (respectively the age at which Antoine stopped his courses) represent Whole 2.

This allowed for variations being introduced between problems, so that the solving strategies would be slightly modified without changing the mathematical structure of the problems too deeply (see Table 1).

Table 1. Composition of the problem statements.

	Cardinal quantities	Ordinal quantities
Common to all problems	A bag of potatoes weighs 5 kilograms. It is weighed with a pumpkin. The weighing scale indicates a total of 11 kilograms. The same pumpkin is weighed with a bag of carrots.	Sophie's travel lasts for 5 hours. Her trip happens during the day. When she arrives, the clock indicates it's 11a.m. Fred leaves at the same time as Sophie did.
V0: identical to the source	The weighing scale indicates 2 kilograms less than before. How much does the bag of carrots weigh?	He arrives 3 hours earlier than she does. How long does Fred's travel last?
V1: inverted operands (question bearing on a whole instead of a part)	The bag of carrots weighs 2 kilograms less than the bag of potatoes. What is the weight indicated by the weighing scale?	His is 3 hours shorter than Sophie's. At what time does Fred arrive?
V2: inverted operator (addition instead of subtraction)	The weighing scale previously indicated 2 kilograms less than it does now. How much does the bag of carrots weigh?	Sophie arrives 3 hours earlier than Fred does. How long does Fred's travel last?
V3: inverted operator and inverted operands	The bag of potatoes weighs 2 kilograms less than the bag of carrots. What is the weight indicated by the weighing scale?	Sophie's travel is 3 hours shorter than Fred's. At what time does Fred arrive?

Hypotheses. First, we hypothesized robustness of encoding effects that would result in poorer transfer of the matching strategy for problems involving quantities that promote cardinal encodings than for problems involving quantities that promote ordinal encodings. More specifically, we tested positive transfer in two different conditions, one where the transfer requires a simple matching of the values, and one where it requires an adaptation of the taught strategy:

- (H1) Even if a literal application of the algorithm leads to success, encoding effects should be observed and thus the matching strategy should be less often transferred if the quantities promote a cardinal encoding than an ordinal one.
- (H2) When the problem test varies with respect to the nature of the question (H2a), or the sign of the difference (H2b) or both (H2c), the participants should show more aptitude to use the matching strategy in the case of an appropriate (ordinal) encoding than if it elicits complementation relations (cardinal encoding).

Second, we investigated the extent to which negative transfer could be observed. We hypothesized that failure to solve the modified problems could mainly be explained by a poor semantical encoding of the examples, leading to non-semantically-based use of the taught algorithm, that is a literal transposition of this algorithm. We thus expected that when the test problems differ from the training problems regarding the nature of the question (H3a), the sign of the difference (H3b) or both (H3c), the errors

made by the participants should vary accordingly, therefore showing that these participants did not properly encode the situation, and were not able to extract the conceptual structures from the training problems.

Methods

Participants. Participants were 110 children ($M = 11.1$ years, $SD = 7.8$ months, from 9.5 to 13.3 years, 5th and 6th grades) who were attending school in the Paris area. They came from 7 different classes in 6 different schools, and came from various socioeconomic backgrounds. They participated voluntarily and ignored our hypotheses.

Design. Each child was presented with a set of problems constituted of 2 training problems and 8 test problems. As mentioned above, the training problems always involved an ordinal quantity. They were duration problems emphasizing the ordinal coding as shown in Gamo et al. (2010). Three bimodal factors were varied across problems: First, the nature of the quantity (cardinal and ordinal). There were 4 types (see Table 2) of quantities: two were cardinal (price and weight); and two were ordinal (distance and temperature). Secondly, the target of the question (part or whole): there were four problems in which the difference between the two wholes was provided and participants had to find the missing part, and four other problems in which the difference between the two parts was given, and subjects had to find the missing whole. Finally, the sign of the difference (+/-): the difference given could either require the subject to perform a subtraction or an addition when using a matching strategy, depending on whether the first or the second of the two elements compared was the highest.

Materials. The problems were printed in booklets. The front page displayed the two training problems which illustrated the matching strategy solution was provided for both of them. The following instructions were given on the upper-side of the page: “You will find an arithmetic problem on every page of this booklet. We ask you to take the time to thoroughly read the problems: there is no time constraint. Please write down every operation you do in order to reach a solution. Just below, you will find two training problems, followed by their respective solutions. Every other problem in this booklet can be solved using the same principle, with only one operation.”

Each test-problem page was divided in three parts: the problem itself was presented on the upper-left side of the page, the response area was on the upper-right side of the page, and an area that could be used as a draft was on the bottom of every test page. These test pages were always presented on the right side of the booklets, while the two training problems with their solution with the matching strategy were displayed on each left page. They were always kept in sight during the test phase, as a reminder.

Procedure. The children were given the booklets and asked to read carefully the front page before starting to solve the problems. After they had answered each of the 8 problems, their booklets were collected. They were told to take all the time they needed, and no participant exceeded the 1 hour time limit we set.

Coding and scoring. A problem was considered as correctly solved when the exact result was found and accompanied by the appropriate calculations. The successful strategies were categorized (correct matching, correct complementation) and so were the incorrect ones (matching with inverted operator, matching with inverted operands, matching with inverted operator and inverted operands, complementation with error, irrelevant, skipped).

For the successes, we used a success-score created to highlight the distribution of matching strategies among the correct strategies: each problem solved successfully with the matching strategy was given a score of 1 and 0 otherwise. We balanced this score using the global success rate for every problem.

For the errors, we calculated 3 error-scores: a matching with inverted operator score, matching with inverted operands score, and a matching with inverted operator and inverted operands score. For each of these scores, we attributed 1 to every congruent error and 0 otherwise. We balanced this score using the global error rate for every problem.

Results

Conditions of positive transfer. We first analyzed, for each problem, the proportion of matching strategies among all the correct trials (see Fig. 2). In order to test our first hypothesis (H1), we examined the frequency of use of the matching strategy on the problems that were identical to the training problems with respect to their mathematical form (same operator, same operands). As hypothesized, the success-score for problems eliciting an ordinal representation ($m = 0.893$, $sd = 0.793$)

was higher than the one for those eliciting a cardinal one ($m = 0.500$, $sd = 0.805$). A paired t-test on the success-scores of these two conditions confirmed that the use of the matching strategy was significantly lower, $p < .001$, when the problems induced a cardinal representation. This result shows that the effect of the type of representation significantly affects the transfer mechanisms; namely, the ordinal problems still lead to a more frequent use of the matching strategy, even though the participants have been encouraged to use the training strategy for all the problems. As for H2, we first considered problems that used the same operator as in the training problems (same sign of the difference) but which required to change the operands, due to the different target of the question. We compared the proportion of matching strategy in cardinal problems with the proportion of this strategy in ordinal problems with a paired t-test on the success-scores. The analysis revealed that there was significantly more matching strategy use in ordinal problems ($m = 0.647$, $sd = 0.932$) than in cardinal problems ($m = 0.182$, $sd = 0.584$), $p < .005$, therefore confirming H2a. This indicates that when the nature of the question changes, the mapping is higher between the training problems and the ordinal test problems, than between the training problems and the cardinal test problems.

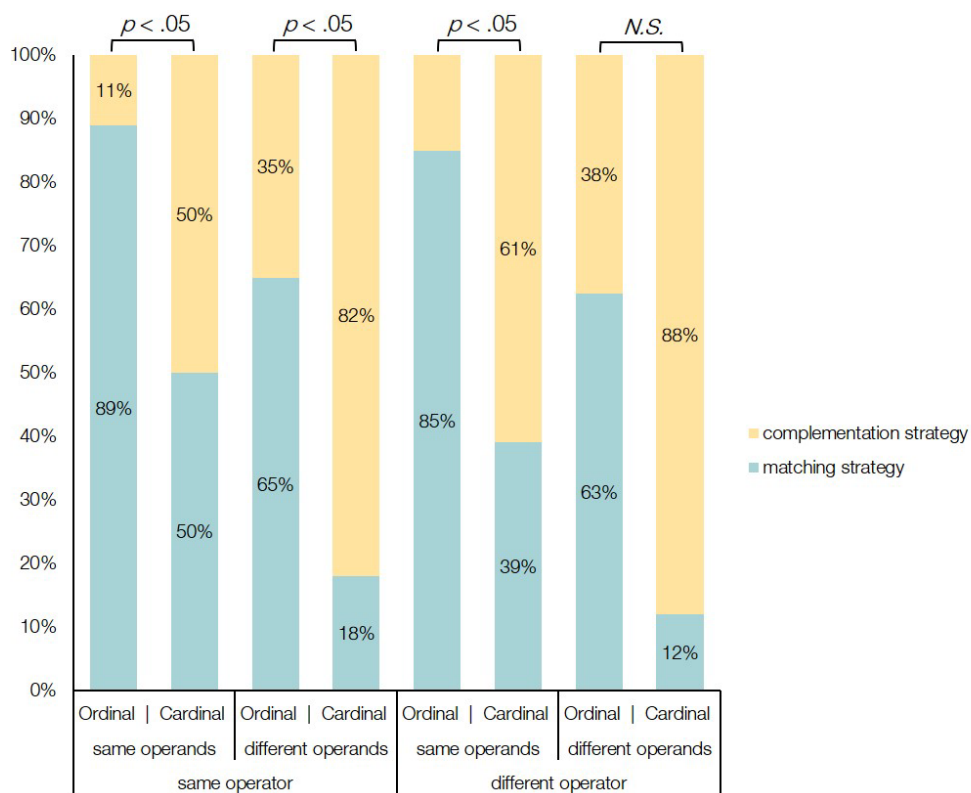


Fig. 2 Proportions of correct solutions by matching strategy and complementation strategy, as a function of the similarity between the training problems and the test-problems. The p-values correspond to significant differences between cardinal and ordinal problems in terms of the proportion of correct matching strategies.

Similarly, we studied the results obtained for problems using the same operands as the training problems, but requiring a different operator (the sign of the difference having been changed). A paired t-test showed that the success-scores for cardinal trials ($M = 0.389$, $SD = 0.905$) were significantly lower than those for ordinal trials ($M = 0.852$, $SD = 1.451$), $p < .01$ in accordance with H2b.

Finally, we studied the use of the matching strategy when both the operands and the operator of the problems differed from those of the training problems. In this case also the paired t-test revealed a significant difference, $p < .05$, between the success-scores of ordinal problems ($M = 0.625$, $SD = 1.503$) and those of cardinal problems ($M = 0.118$, $SD = 0.676$). The problems inducing an ordinal representation therefore seem to facilitate the use of the strategy learnt, even when it requires adapting two different factors in order to be used, in conformity with H2c.

Thus, the influence of the semantic encoding of the problems has a strong impact on transfer. It is so robust that even in the case of additive problems in which a source is repeatedly shown to the participants with a solving strategy in one operation that leads to the solution, and that it is explicitly stated in the instructions that the same solution in one operation applies to all the problems, participants mostly use the longer three steps strategy when the quantities involved promote a cardinal encoding. In contrast, most of them use the one step strategy when the quantities involved promote an ordinal encoding. This holds true both when a literal application of the taught algorithm is enough (H1) and when this taught algorithm has to be adapted (H2).

Analysis of negative transfer. The second part of our analysis involves the study of the distribution of errors across the experimental conditions. We created the following typology for the “matching with error” strategies used by participants:

- (i) correct operands with the wrong operation (an addition when a subtraction is needed, or conversely), were classified as an “inverted operator only” error.
- (ii) correct operator with the wrong values (calculating the whole when the question is about the part, or conversely), was classified as an “inverted operands only” error.
- (iii) wrong operation and wrong values, was an “inverted operator and inverted operands” error.
- (iv) all the other errors (use of multiplication or division, use of more than one operation leading to an incorrect result, absence of use of the difference value,

use of a complementation strategy leading to a failure), were labeled “other errors”.

In the following analyses, we grouped together the problems that were promoting ordinal or cardinal encodings, because there was no specific prediction and paired t-tests showed no significant differences. We performed the following analyses:

We first analyzed how the ‘inverted operands only’ errors were distributed across the different types of problems (see Fig. 3, 1st bar). We compared test problems which were identical to the training problems (same sign of the operator and same operands), with problems in which only the choice of the operands differed from the training problem. We compared the error-scores for ‘inverted operator only’ strategies in these two conditions with a paired t-test. The results showed that the proportion of errors in the case of problems with inverted operands ($M = 0.629$, $SD = 0.959$) was significantly higher than the proportion of errors for problems identical to the training problems ($M = 0.229$, $SD = 0.605$), therefore supporting H3a, $p < .05$.

Regarding the ‘inverted operator’ errors (see Fig. 3, 2nd bar), similarly, we compared the test trials which were identical to the training examples, with the test problems which differed in terms of the operator (i.e., to perform an addition rather than a subtraction). A paired t-test analysis showed that problems with an inverted operator ($M = 0.777$, $SD = 1.174$) gave significantly more ‘inverted operator’ errors than the problems with no change from the training ones ($M = 0.112$, $SD = 0.540$), $p < .001$, therefore supporting H3b.

Finally, we compared the proportion of ‘inverted operands and inverted operator’ errors (see Fig. 3, 3rd bar) in problems homologous to the training problem and problems in which both the operator and the operands were inverted. The no-change condition showed significantly less errors of this category (0.107 , $SD = 0.724$) than for problems with both an inverted operator and inverted operands (0.760 , $SD = 1.274$), $p < .001$, thus supporting H3c.

Overall, these results suggest that participants who failed to encode the problems in an appropriate manner (either through a cardinal or an ordinal encoding) and failed to solve the problem, were influenced by the taught algorithm but applied it in a literal way. Indeed, errors resulted from a literal transposition of the calculations provided in the example problem.

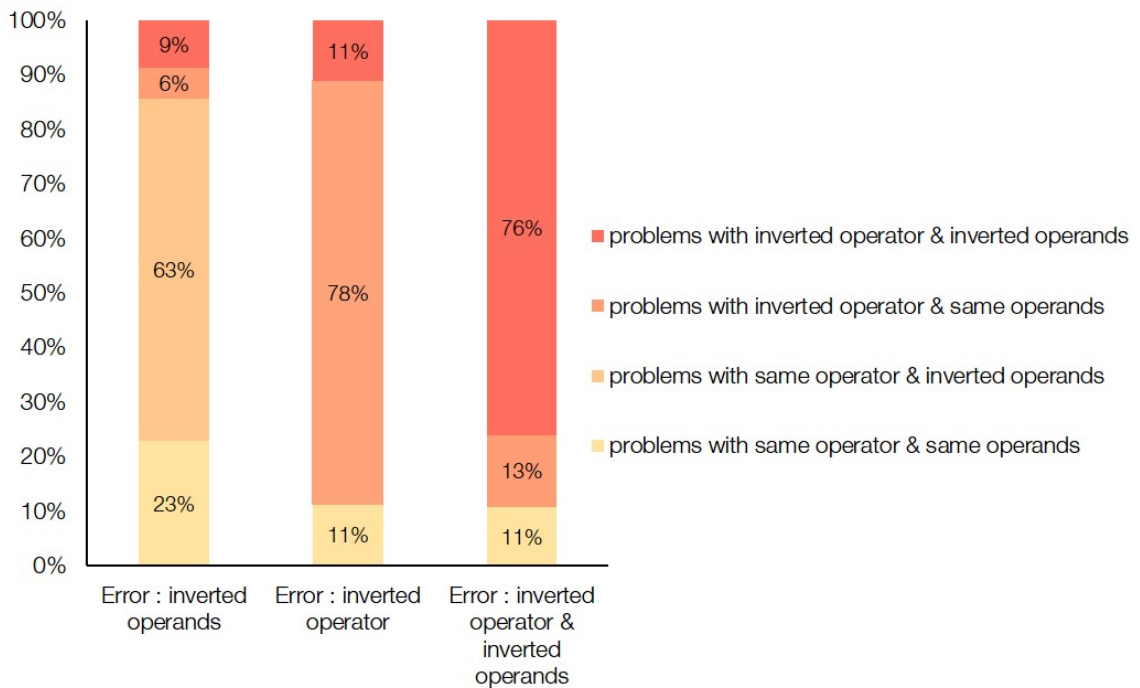


Fig. 3 Distribution of the different type of errors across the problems.

Discussion

In accordance with the hypotheses, when the quantities involved in the problems were promoting a cardinal encoding, a large proportion of participants failed to apply the algorithm they were taught to novel examples both when they were sharing all the characteristics of the training problems and when they differed in terms of the operands and/or the operator. This suggests that the representations induced by “what we know about the world” was not compensated by the explicit teaching of the matching strategy and the explicit instruction to use it. The fact that this effect persisted even when the use of the matching strategy was made less obvious by the modifications introduced between the training and the test acknowledges the importance of this effect.

Recent work (e.g. DeWolf, Bassok, & Holyoak, in press; Rapp, Bassok, DeWolf, & Holyoak, in press) emphasizes the generality of the phenomenon of semantic alignment and the underlying educational perspectives.

In this work we showed that initial spontaneous encoding is highly influential in that it does not only constrain the strategy spontaneously used as it was shown by Gamo et al. (2010), but it also impairs transfer even in situations in which the solution provided required low technical knowledge (additions and subtractions) and relied

conceptually on simple mathematical relations (looking for a part or a whole or comparing).

This phenomenon highlights the importance of overcoming the initial encoding in some cases, even when this initial encoding is relevant from a mathematical point of view: the cardinal encoding and the derived complementation strategy are relevant for solving the problems in this study but have to be overcome in order to apply successfully the matching strategy. A general encoding such as the one symbolized in Fig. 1 is far from being spontaneous. This is a promising and challenging route to develop methods for semantic recoding that are more general than the initial one in that they remain compatible with it but embrace a larger number of situations and are mathematically more apt.

References

- Bassok, M, Chase, V. M, & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology*, 35, 99-134
- Bassok, M., & Olseth, K., (1995). Object-based representations: Transfer between cases of continuous and discrete models of change. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1522-1538.
- Bassok, M., Wu, L., & Olseth, L. K. (1995). Judging a book by its cover: interpretative effects of content on problem-solving transfer. *Memory and Cognition*, 23, 354 e 367.
- Brissiaud, R., & Sander, E. (2010). Arithmetic word problem solving: a Situation Strategy First Framework. *Developmental Science*, 13, 92-107.
- 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.
- DeCorte, 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.
- DeWolf, M., Bassok, M., & Holyoak, K.J. (in press). Conceptual structure and the procedural affordances of rational numbers: Relational reasoning with fractions and decimals, *Journal of Experimental Psychology: General*.
- Gamo, S., Sander, E., & Richard, J.-F. (2010). Transfer of strategy use by semantic recoding in arithmetic problem solving. *Learning and Instruction*, 20, 400-410.
- Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 398 e 415.
- Rapp, M., Bassok, M., **DeWolf, M.**, & Holyoak, K. J. (in press). Modeling discrete and continuous entities with fractions and decimals. *Journal of Experimental Psychology: Applied*.
- Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem solving ability in arithmetic. In H.P. Ginsburg (Ed.), *The development of mathematical thinking*. New York: Academic Press.

- Ross, B.H. (1989). Distinguishing types of superficial similarities: different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 456-468.
- Thevenot, C., & Oakhill, J. (2006). Representations and strategies for solving dynamic and static arithmetic word problem: The role of working memory capacities. *Quarterly Journal of Experimental Psychology*, 58, 1311-1323.
- Vicente, S., Orrantia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology*, 77, 829-848.

GENERAL DISCUSSION



Our contribution

At the beginning of this thesis, we set out to understand the extent to which human reasoning is constrained by the content on which it operates. Namely, we wanted to show that general, non-mathematical knowledge could influence the encoding of arithmetic word problems by highlighting either the cardinal or the ordinal property of the numbers they feature. We predicted that this difference in encoding would interfere with participants' understanding of the problems and subsequently shape their behavior in a selection of tasks. The first chapter described a general framework outlining the respective roles that world and mathematical semantics may play in the interpretation of arithmetic word problems. The take-home message of this chapter mainly consists in the definition of semantic (in)congruence, the notion that guided our exploration of cardinality and ordinality in arithmetic word problems. Across the following 6 chapters and 16 experiments, we sought to establish the relevance of the distinction between cardinal and ordinal problems, assessing the extent of its influence on a wide range of activities. At this point, a brief overview of what can be concluded from our work – and, by contrast, of what remains to be elucidated – seems in order.

Key findings

The finding that was the most replicated between the experiments is the fact that participants tend to use different algorithmic strategies to solve isomorphic problems, depending on the quantities they feature. When two algorithms could be used, finding the shorter one was easier on ordinal problems than on cardinal problems⁴. This imbalance in strategy use was our starting point, since Gamo, Sander, and Richard's (2010) results suggested that problems counting family members, item prices and event durations may be solved differently by 5th graders. Here, we replicated this finding with other quantities (number of floors, temperatures, heights, weights, prices, durations, collections), other populations and different instructions. Similarly, when only the 1-step algorithm was available, participants still tended to use it more often on ordinal than on cardinal problems.

⁴ The only exception being the adults who solved the simplified problems in Chapter 6 (Experiment 2). They did, however, use the 3-step algorithm more often on cardinal problems than on ordinal problems.

Another important finding of this thesis regards the fact that those performance differences can be traced to differences in the encoded representations of the problems. With the help of drawing tasks, recall tasks, classification tasks, comparison tasks, sentence recognition tasks, and eye movement recordings, we were able to probe participants' representation of cardinal and ordinal problems and to highlight crucial differences between them.

Another central result of our experiments resides in the difficulties displayed by participants when trying to overcome semantic incongruence. Error rates, response times, negative transfer and pupil dilation all showed that using the 1-step algorithm on cardinal problems involved a cognitively demanding process. While the precise nature of such a process is still up for debate, its existence was supported by several converging measures.

Finally, we believe that the prevalence of these effects of semantic (in)congruence has been substantially supported by the diverse populations recruited in the experiments. From 2nd graders to mathematicians, also including lay adults, 5th graders, and math teachers, they all seemed influenced by the cardinal-ordinal distinction.

Issues requiring further investigation

While some of our questions have been answered within this thesis, it should also be noted that others have been raised over the experiments we conducted. Among them, an important issue regards the process by which one overcomes a difficulty stemming from semantic incongruence, adopting a new point of view to find a solution. We have called this process semantic recoding and, as previously stated, we have gathered evidence of its existence, but its inner mechanics remain to be elucidated. Does semantic recoding rely exclusively on mathematical semantics, or do individuals also use world semantics when constructing a new representation? Is this process composed of one single recoding, a sudden reconceptualization akin to the '*Aha!*' moment described in the literature on insight (Kounios & Beeman, 2009), or are several successive recodings involved? What happens when a new representation is constructed that does not lead to any new solving strategy? When a problem admits more than two relevant encodings, is it possible to recode an interpreted structure into another interpreted structure without encoding the problem's deep structure?

These questions remain open and call for further investigation. For instance, the use of new problem statements allowing for more than 2 different solving

algorithms may provide valuable insights regarding the successive problem representations constructed by the solvers. Another possible route to better understand this process may lie in the use of electrophysiological responses such as event-related potentials. Indeed, the N400 – a negative-going voltage shift that peaks around 400ms after the onset of a stimulus – is thought to be associated with conceptual integration disruptions (Kutas & Federmeier, 2000; Osterhout, Holcomb, & Swinney, 1994). In other words, its amplitude is generally assumed to be modulated by the degree to which a conceptual or grammatical violation disrupts integration. Guthormsen et al. (2016) showed that arithmetic problems that are semantically misaligned elicit stronger N400 effects than semantically aligned problems. Thus, ERP recording may provide insights into what semantic incongruence entails for the solvers, in terms of conceptual integration, and help characterize the ensuing recoding process. Similarly, thinking aloud protocols – where participants are asked to comment their reasoning while they solve the problems – might also be an interesting lead to get a better grasp of the semantic recoding process. Furthermore, another path to investigate semantic recoding may reside in adjusting the cognitive load associated with a solving algorithm. Indeed, Thevenot and Oakhill (2005, 2006) showed that participants could be influenced to change the strategy they used to solve a problem simply by increasing its cost in working memory. For instance, the authors showed that increasing a problem's numerical values from 2-digit numbers to 3-digit numbers led participants to construct an alternative mental representation of the problem and use a different solving algorithm. Similarly, using higher values on cardinal problems may encourage participants to construct a new representation and use the 1-step algorithm requiring fewer calculations.

Another point worthy of attention in future works regards the potential generalization of our findings. Although we strove to provide exact and conceptual replication among the different experiments, our materials remained relatively similar overall. Our claim could thus benefit from new experiments conducted with different materials and by different research teams. In view of the current replicability crisis in psychology, researchers advocate for systematic replication of findings by independent labs (Open Science Collaboration, 2015; Simons, 2014; Stroebe & Strack, 2014). On this topic, we are currently engaged in a collaboration with two other teams in order to evaluate how cardinality and ordinality interact with other problem attributes such as familiarity of the quantities and cognitive load. For instance, recently

collected preliminary data seem to indicate that transforming a “number of floors” ordinal problem into a “number of tree branches” problem – in which the elevator is replaced by an animal going up a tree by a certain number of branches – does not significantly change participants’ ability to use the 1-step algorithm. Similarly, apart from durations, heights, prices, weights, temperatures, collections and number of floors, what other quantities can be used to foster one of two encodings? Can we replicate the effect of cardinal and ordinal quantities on problems with a different mathematical structure? Addressing these questions in future studies would likely help ascertain the generalization of our findings.

Finally, we believe that the SECO model would benefit from additional empirical support. Namely, experiments specifically targeting its predictions are necessary to assess its explanatory power. As previously stated, the experiments we reported in this thesis did not specifically aim at assessing the validity of SECO’s predictions. We believe their results support SECO’s general claim – that world semantics plays a non-negligible part in the encoding, recoding and solving of arithmetic word problems – but the finer details of the model could benefit from further investigation. For instance, can we find evidence for the existence of an “expert encoding” pathway? Can we empirically demonstrate that an individual may construct different, subsequent representations while solving a problem? May participants with different world semantics encode different interpreted structures from the same problem statement? We hope to investigate these exciting questions in further studies.

Before concluding this thesis, we propose to tackle a few related issues that may shed some light on our interpretation of its findings, dissipate potential misunderstandings and help develop new empirical prospects.

The paradox of P-items?

Throughout this thesis, we have striven to show that daily-life, non-mathematical knowledge about the world interferes with arithmetic reasoning, notably among children. Yet, at first glance, these findings may seem paradoxical to the many empirical reports stating that children tend to disregard what they know about the world when trying to solve arithmetic word problems.

The most famous example of children giving an apparently nonsensical answer to an arithmetic problem probably comes from Baruk (1985), who created problems

such as “There are 26 sheep and 10 goats on a ship. How old is the captain?”. Baruk found that children overwhelmingly replied to such problems by using the provided numerical values, while only a minority of participants pointed out the inadequacy of the question and the impossibility to find the answer. The idea that children could manifest a “suspension of sense-making” (Verschaffel, Van Dooren, Greer, & Mukhopadhyay, 2010, p.12) and exclude real-world knowledge from their responses to mathematical word problems was further investigated through the creation of “problematic problems” (called P-items) requiring the use of common sense to be solved (Greer, 1993; Verschaffel, De Corte, & Lasure, 1994). Contrary to the captain’s problem, those P-items admitted sensible answers, although unusual ones that would not normally be expected in a mathematical classroom. For instance, the problem “Steve has bought 4 planks each 2.5 meters long. How many 1-meter-long planks can he saw from these planks?” required to realize that two 0.5-meter planks did not make a 1-meter plank. Replying with the algorithm $4 \times 2.5 = 10$ was deemed a “non-realistic reaction”, whereas a “realistic reaction” was either to provide the correct answer ($4 \times 2 = 8$) or to reply that the solution was not straightforward and acknowledge the unrealistic nature of the 4×2.5 algorithm. Greer (1993) and Verschaffel et al. (1994) showed that the vast majority of children tended to display non-realistic reactions to P-items, whereas they had realistic reactions to non-problematic control problems. In other words, students tended to ignore real-world knowledge as well as realistic considerations from the answers they gave: they replied by applying the mathematical procedures they learned, regardless of their relevance in the described situation. Since then, this effect has been replicated on numerous occasions, in several countries (Caldwell, 1995; Csikos, 2003; Hidalgo, 1997; Renkl, 1999; Reusser & Stebler, 1997; Verschaffel, et al., 1999; Xin, Lin, Zhang, & Yan, 2007; Yoshida, Verschaffel, & De Corte, 1997; see Verschaffel et al., 2010, for a review).

How can we, then, account for the fact that children seem to disregard real-world knowledge when finding the solution to a mathematical word problem, while at the same time arguing that children and adults alike are influenced by their world knowledge when solving cardinal and ordinal problems? We believe that the existence of these two types of effects is not contradictory. In fact, while it is indisputable that children give unrealistic answers due to them putting aside real-world knowledge when solving mathematical word problems, this might have more to do with their representation of what is expected of them, rather than with an

inability to come up with realistic solutions. In this regard, DeFranco and Curcio (1997) compared students' performance on P-items depending on whether the problems were presented in a traditional scholastic setting, or in a more ecological setting, closer to a real-life situation. They asked students to solve the following problem "328 senior citizens are going on a trip. A bus can seat 40 people. How many buses are needed so that all the senior citizens can go on the trip?" Only 10% of the students managed to find the solution to this problem. In a second part, they asked the same children to make a call to book minivans for a school trip, with a similar problem regarding the number of minivans. In this more realistic setting, 80% of the students booked the appropriate number of minivans, despite most of them having failed in the first task. Thus, children's ability to think realistically was directly dependent on the context in which the problem was presented. They suspended their common sense when faced with a math problem, but not when asked to make a "real" phone call to book the appropriate number of vehicles. This result is reminiscent of the effect reported in the introduction of this thesis, regarding the increase in performance in the "realistic version" of the Wason task, where participants had to make sure that no underage client was drinking alcohol. Overall, it thus seems that the problem with P-items does not come from an inability to use any form of real-world knowledge to evaluate the relevance of a mathematical procedure.

In contrast, we believe that the results that have been reported in this thesis have to do with a different, more insidious effect than that underlined by P-items. We believe that the difference between cardinal and ordinal problems does not have to do with realistic considerations that participants would fail to take into account. Rather, the world semantics imbued in the problems influenced the very encoding of the problems, because of the habits we developed of representing durations along a timeline or collections as unordered sets. At no point do pragmatic considerations regarding how realistic the solutions are factor in the participants' reasoning. The fact that the difference between cardinal and ordinal problems was not only found in children, but also among lay adults, math teachers and expert mathematicians should be enough to consider that we are dealing with two independent phenomena. Children do tend to ignore real-world knowledge when considering how realistic the solution to a P-item is, but children and adults also tend to be influenced by their real-world knowledge in their initial encoding of a problem's situation.

Beyond the dichotomy

Despite this manuscript focusing on the differences between cardinal and ordinal problems, it should be noted that our claim is not that there exists a strict dichotomy between problem statements systematically fostering a cardinal encoding and problem statements systematically evoking an ordinal encoding. Rather, we believe that a more accurate account would be to consider that there is a gradient of cardinality and ordinality, on which different problem statements can be placed, depending on how strongly they emphasize the cardinal or the ordinal nature of the numbers they use. For example, the collection problems we created seem to be more prototypical of cardinal problems than the price problems, meaning that it seems easier to discard a cardinal encoding and construct an ordinal encoding based on a situation in which prices are calculated, as compared to a situation where, for instance, marbles are counted. Similarly, participants seem to have a slightly easier time using the 1-step algorithm on elevator problems than on height problems, which suggests that elevator problems are more likely to elicit an ordinal encoding of the situation. In fact, apart from the simplified problems used with 2nd graders in Chapter 6 (Experiment 2), none of the problems used in the other experiments were ever solved using only one of the two algorithms by every participant. When the problems could be solved with both algorithms, there was usually at least one participant to use the 1-step algorithm, and one to use the 3-step algorithm. What differed between cardinal and ordinal problems was the proportion of participants who used each of these two algorithms.

The analyses of participants' strategies to solve hybrid problems, in Chapter 2 (Experiment 3), demonstrated that the use of cardinal quantities in a problem statement does not systematically mean that it will be encoded into a cardinal representation. By presenting cardinal quantities fluctuating across time (e.g. animal gaining weight over the years), we were able to facilitate the use of the 1-step algorithm, albeit participants were still less likely to use it than when solving a problem that only featured ordinal quantities. We believe that the existence of such "hybrid" problems that can foster either an ordinal encoding or a cardinal encoding indicates that the same situation can be interpreted differently by different solvers or by the same solver over time. This raises two questions. One has to do with transfer and the role that hybrid problems may play in its promotion. It will be discussed later

in the discussion. The other one has to do with the variations that may occur between individuals, depending on their prior knowledge and cultural background.

Indeed, since our view is that world semantics has an influence on one's encoding of a problem statement into an interpreted structure, then it follows that depending on one's state of knowledge about the entities described in a problem statement, the interpreted structure may vary. Does extensive experience with elevators increase our ability to see number of floors problems as ordinal? This question, worth considering, raises another: how does one's cultural setting factor into their interpretation of mathematical word problems? Regardless of the intercultural differences existing in individuals' numerical cognition (e.g. Dehaene, Izard, Spelke, & Pica, 2008; Pica, Lemer, Izard, & Dehaene, 2004; Reys, et al., 1999), do cross-cultural differences in non-mathematical cognition influence mathematical word problem solving? Or, in other terms, does the way we speak about the world influence the way we count its entities? In the next section, we discuss how the materials created in this thesis may offer a path to investigate this issue.

Taking a turn for the Whorf

The idea that the language we speak may affect the way we think about the world is not new. Ever since Benjamin Whorf (1939) formulated what would later be known as the Sapir-Whorf hypothesis of linguistic determinism, the notion that human language constrains cognition has been a heated topic among linguists. Despite its initial popularity during the first half of the 20th century, linguistic relativism rapidly fell out of favor among cognitive scientists and linguists. Notably, studies suggesting the existence of universal semantic constraints on color terminology (Berlin & Kay, 1969), works falsifying some of Whorf's most famous claims on Eskimo lexicon (Pullum, 1991), and researchers pointing out the universality of non-linguistic concepts (Chomsky, 1975; Fodor, 1975; Pinker, 1994) all contributed to marginalize the Sapir-Whorf hypothesis and its tenants.

Only recently has linguistic relativism resurfaced, in its “weaker version” stating that language has a non-deterministic influence on cognitive processes. For instance, Oh (2003) showed that Korean and English speakers differed in their recall of motion events, due to motions being encoded in terms of “paths” or “manner” of motion depending on the language. Similarly, Boroditsky (2001) showed that Mandarin speakers were faster to judge sentences about temporal succession (e.g.

what is the year before the year of the tiger?”) when primed with a vertical spatial stimulus, whereas English speakers were faster when primed with a horizontal spatial event, due to time being generally represented horizontally in English and vertically in Mandarin. Despite showing significant influence of one’s native language on memory and time perception, the results of these studies have been somewhat downplayed as a case of “thinking for speaking” (Slobin, 1996). Since, in both studies, participants were thinking with the intent to use language to describe the events, it might be that the effects observed were only the result of their attempt to parse the events in linguistic terms, and thus that language does not necessarily influence cognition when the expected response is non-linguistic (Papafragou, Massey, & Gleitman, 2002). Thus, researchers have attempted to determine if crosslinguistic differences may have consequences such that people who use different languages end up thinking differently, even in situations where no linguistic response is expected. As Casasanto (2008, p.69) puts it in his article “Who’s afraid of the big bad Whorf?”, the question remains open and “one obstacle to resolving this controversy has been devising truly nonlinguistic tests to evaluate how speakers of different languages perceive or remember their experiences”.

In this perspective, an ingenious experimental paradigm was developed by Casasanto, Fotakopoulou and Boroditsky (2010) to study how language influences our non-linguistic time representation. In their experiments, participants watched lines “growing” across a screen, over a certain time. They were asked to estimate the duration for which the lines grew, regardless of how long the line itself was, by clicking the mouse for an equal amount of time. This paradigm makes it possible to evaluate how distance perception interferes with line perception on a physiological level. Casasanto et al. (2010) predicted that speakers of different languages would not be influenced to the same extent by the growing lines, due to the different words used to describe time in their respective language. Interestingly, Casasanto et al. (2004) had underlined that in English and in Indonesian, the most commonly used expressions to describe durations are distance related (e.g. “a long time” in English), whereas in Spanish and in Greek, the most used expressions are quantity related (e.g. “mucho tiempo” in Spanish). They hypothesized that this lexicon difference would imply a difference in time representation. In other words, languages such as English and Indonesian seem to promote an axis-based representation of time, due to the use of linear spatial expressions to describe durations, whereas Greek and Spanish seem

to promote a set-based representation of time, due to the use of terms describing amounts instead of distances (e.g. “poli”, in Greek, or “mucho”, in Spanish). Casasanto et al. (2004) confirmed this view by showing that English and Indonesian speakers’ perception of time was strongly influenced by the total distance traveled by the growing lines, whereas no such effect appeared for the Spanish or Greek speakers.

It will not have escaped the attentive reader that the distinction between the representation of time in languages using linear spatial expressions to describe durations (“long time”, in English, “waktu panjang”, in Indonesian, “longtemps” in French) and that of languages describing time using amounts (“mucho tiempo” in Spanish, “poli ora”, in Greek) is remarkably close to the distinction, introduced in this thesis, between cardinal and ordinal representations. It follows directly from this observation that speakers of different languages may behave differently when attempting to solve duration problems. If in fact, Spanish and Greek speakers have a more cardinal conception of time, due to the vocabulary they use, then they may be more likely to use the 3-step algorithm to solve such problems, as compared to French or English speakers. Similarly, would they have a harder time assessing the solvability of duration problems, in experiments such as the ones described in Chapters 4 and 5? We believe that this hypothesis constitutes a promising prospect for investigating the influence of language on non-linguistic reasoning. To the extent of our knowledge, there have been no attempts to assess the validity of the Whorfian hypothesis using arithmetic word problems. Even though, by definition, arithmetic word problems are stated using words, the expected response is a mathematical one, not a linguistic one. Thus, it can be argued that using our duration problems to investigate cross-linguistic differences in time representation would not fall within the range of “thinking for speaking” tasks. By comparing, for instance, Spanish and English speaker’s ability to use the 1-step algorithm on duration problems and on collection problems (as a baseline), we may be able to show that the language we speak influences the way we count...for better or for Whorf.

How can transfer be improved?

An important part of this thesis has been dedicated to showing the robust influence that our knowledge about the world has on our mathematical reasoning. Regardless of the instructions given to them and regardless of their expertise, participants systematically struggled to use, on cardinal problems, the strategies that they

developed on ordinal problems. The pervasiveness of these worldly interferences raises the question of what it takes to overcome semantic incongruence and reach the same level of performance regardless of the semantics imbued in a problem. How can we help solvers engage in semantic recoding? What routes do we have at our disposal to promote transfer between semantically incongruent situations? Gamo et al. (2010; see also Gamo, Nogry, & Sander, 2014) showed that an extensive training session, led by a teacher, focusing on the deep structure of the problems and the explicit learning of both solving algorithms could help increase transfer to semantically incongruent problems. But such training is costly and does not necessarily entail that participants will be able to engage in far transfer when training and test problems are too dissimilar. Is there no other way to help solvers learn a semantically incongruent algorithm? In Chapter 7, we showed that simply providing participants with a worked-out example was not enough to help them understand how to solve cardinal problems. Although the evaluation of training methods promoting semantic recoding was not directly within the scope of this thesis, we can nevertheless speculate about the merits of alternative approaches to promote transfer.

Hybrid problems as pivotal situations?

As previously mentioned, we believe that hybrid problems such as those created in Chapter – Experiment 4 may play a part in the promotion of semantic recoding and transfer. Indeed, by creating situations in which cardinal quantities evolved along an ordinal axis, we “ordinalized” the cardinal problems, which resulted in a higher chance to find the 1-step algorithm on such problems. Thus, these problems may constitute an entry point to the idea that cardinal problems can be represented along an ordinal axis and solved using the 1-step algorithm. Can we improve transfer of the 1-step algorithm from cardinal problems to ordinal problems by training on hybrid problems first? Thanks to their “hybrid” status, such problems may hold a pivotal role in promoting semantic recoding: if a solver is taught how to see both possible encodings on a single situation, then they may be able to learn how to switch from one to the other. In other words, we believe that hybrid problems may constitute a fertile ground to train switching between two possible encodings of a given situation. Similar to how bistable illusions tend to lead to an alternation between two distinct percepts (Eagleman, 2001), hybrid problems may facilitate alternating between two problem representations. And, just like with bistable illusions, it seems reasonable to expect that some degree of control over (perceptual) reversals may be learnable.

Preliminary data we recently collected may shed some light on this issue. In an attempt to evaluate whether training could improve solvers' ability to see past their initial encoding of a situation, we recruited 138 pre-service teachers and asked them to participate to a 4-step training program spanning over 8 weeks. Two weeks after an initial pre-test on cardinal and ordinal problems, participants were randomly assigned to one of four groups. Each group followed 2 short online training sessions over a 4 weeks period. The sessions involved either cardinal problems, ordinal problems, hybrid problems, or unrelated distributivity problems (control group). In the training sessions, participants were told that a 1-step algorithm could be found to solve the problems, and a series of questions were posed to guide their exploration of the problems. Two weeks after the second training session, participants were asked to complete a post-test. Post-test results are graphically summarized in Fig. 1.

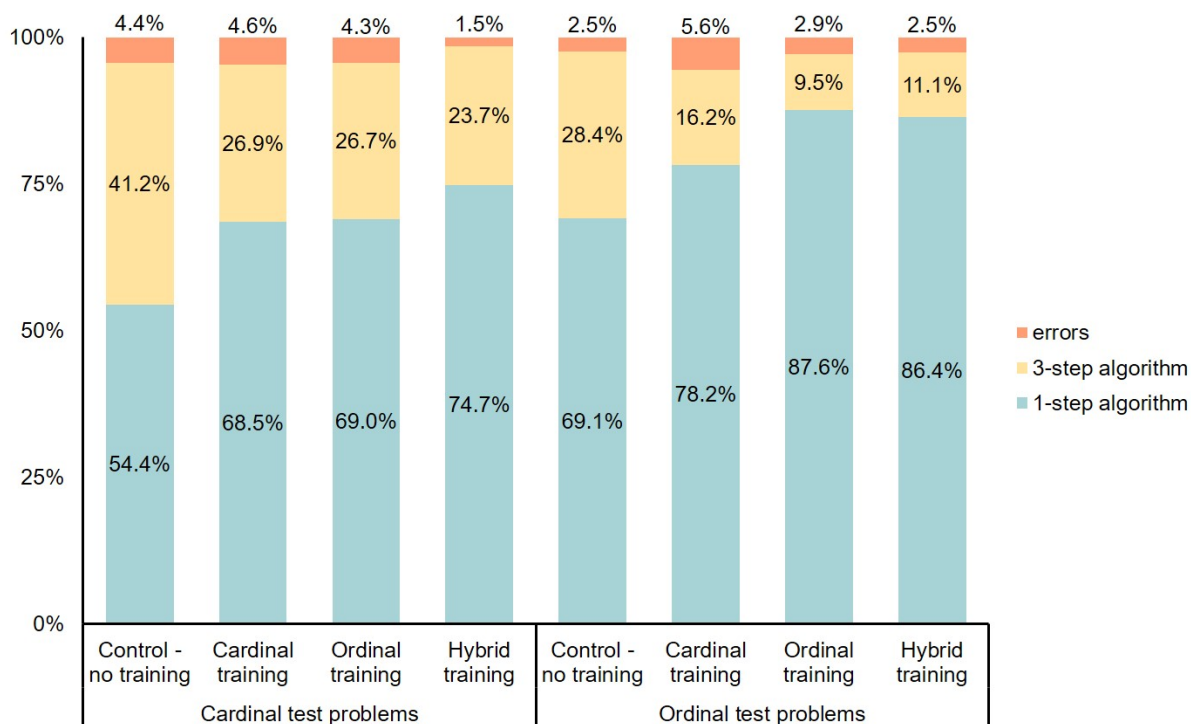


Fig. 1 Post-test performance on cardinal and ordinal test problems, depending on the received training.

Although one should exercise caution in interpreting this graph, as it describes preliminary results, it is interesting to note that the best scores on cardinal test problems were obtained by the group who trained on hybrid problems, who found the 1-step algorithm in 74.7% of the cases. As a comparison, the groups who trained on cardinal or ordinal problems only achieved, respectively, 68.5% and 69% success

on cardinal problems. On ordinal problems however, the highest performances were recorded in the ordinal training group, whose participants used the 1-step algorithm in 87.6% of the cases. The performances of participants in the hybrid training group were not too far behind, as they reached 86.4% use of the 1-step algorithm.

Although more analyses are needed and no definite conclusion should be drawn from these preliminary results, they nevertheless seem to suggest that training on hybrid problems may provide a slight advantage to learn how to engage in semantic recoding on semantically incongruent problems. Thus, we believe that studying hybrid problems for which distinct encodings may be constructed could be a promising route to help promote semantic recoding. Further research may provide valuable insights regarding this issue.

Argumentation to promote semantic recoding?

Another interesting perspective to help promote semantic recoding and transfer in a scholastic setting comes from recent studies on the role of argumentation in the spread of counter-intuitive beliefs. According to the argumentative theory of reasoning, reasoners are more objective and less biased when they evaluate arguments produced by others (Mercier & Sperber, 2011). Thus, this theory predicts that reasoners are more likely to change their view when arguing with other reasoners, than when they are simply presented with statements challenging their view (Mercier, 2016). It is believed that argumentation may hold a special role in helping one change their view and adopt a counter-intuitive perspective. For instance, Claidière, Trouche, and Mercier (2017) conducted an experiment in which they asked groups of participants to solve problems with a counter-intuitive solution, such as the *bat and ball* problem, in either a “discussion” condition or in a “silence” condition. After an initial “individual phase” in which participants were asked to solve a problem by themselves, the participants were allowed to communicate with each other. The groups in the “discussion” condition were told they could discuss their answers with their immediate neighbors, whereas the groups in the “silence” condition were only allowed to look at their neighbors’ answers, without the ability to talk and argue about the problem. Participants were asked to write down their current answers at regular time intervals, to identify if, and when, they changed their mind about the problem’s answer. Results showed that in the “discussion” condition, the majority of the groups ended up with a global consensus on the correct answer, whereas in the “silence” condition, participants who found the correct answer only spread it to a limited number of

individuals, with most participants still set on their initial wrong answer. The authors concluded that argumentation was a powerful tool to help students acquire counterintuitive concepts when other methods fail.

Since our results suggest that recoding the initial representation of a problem is an arduous task, even when participants are explicitly invited to do so, we considered the use of argumentation as a means to overcome semantic incongruence. In an ongoing study conducted with our colleague G ery Marcoux, we asked groups of pre-service teachers to find the shortest possible algorithm to a series of cardinal problems. We sat the participants in a classroom and reproduced Claidi ere et al.'s (2017) "discussion" condition by telling them that they could discuss their answer with their neighbors. The results of one group are reproduced in Fig. 2. After a short individual phase in which participants had to solve the problems by themselves (t_0), they were told that they could start talking with their neighbors and they had to report, after each minute (t_1 to t_{15}), what their current answer was, as well as their degree of confidence in it being the best answer. After the initial individual solving phase, only 3 participants had found the 1-step algorithm, whereas 17 had found the 3-step algorithm. Once participants were allowed to discuss their findings, they rapidly convinced each other of the validity of the 1-step algorithm, and after 9 minutes, all of them agreed that it was the shortest possible algorithm to solve the problem.

We then presented the same group with another cardinal problem (see Fig. 3). After the individual phase, 15 participants had managed to find the 1-step algorithm by themselves! After being allowed to talk for 1 minute, 7 more participants agreed on a 1-step solution as well. After 2 more minutes (t_3), all of the participants were convinced of the validity of the 1-step algorithm they found, with a high confidence rate on average. Thus, it seems that the opportunity to argue over the 1st cardinal problems provided participants with a deep enough understanding of the problem's structure that they were quickly able to transfer the newly learned strategy to a new cardinal problem.

Although these results are only preliminary, as only 4 groups have been recruited at this time and the addition of a "silence" group would allow a finer understanding of the part played by argumentation, we believe that they nevertheless open a promising route to promote semantic recoding in a scholastic setting.

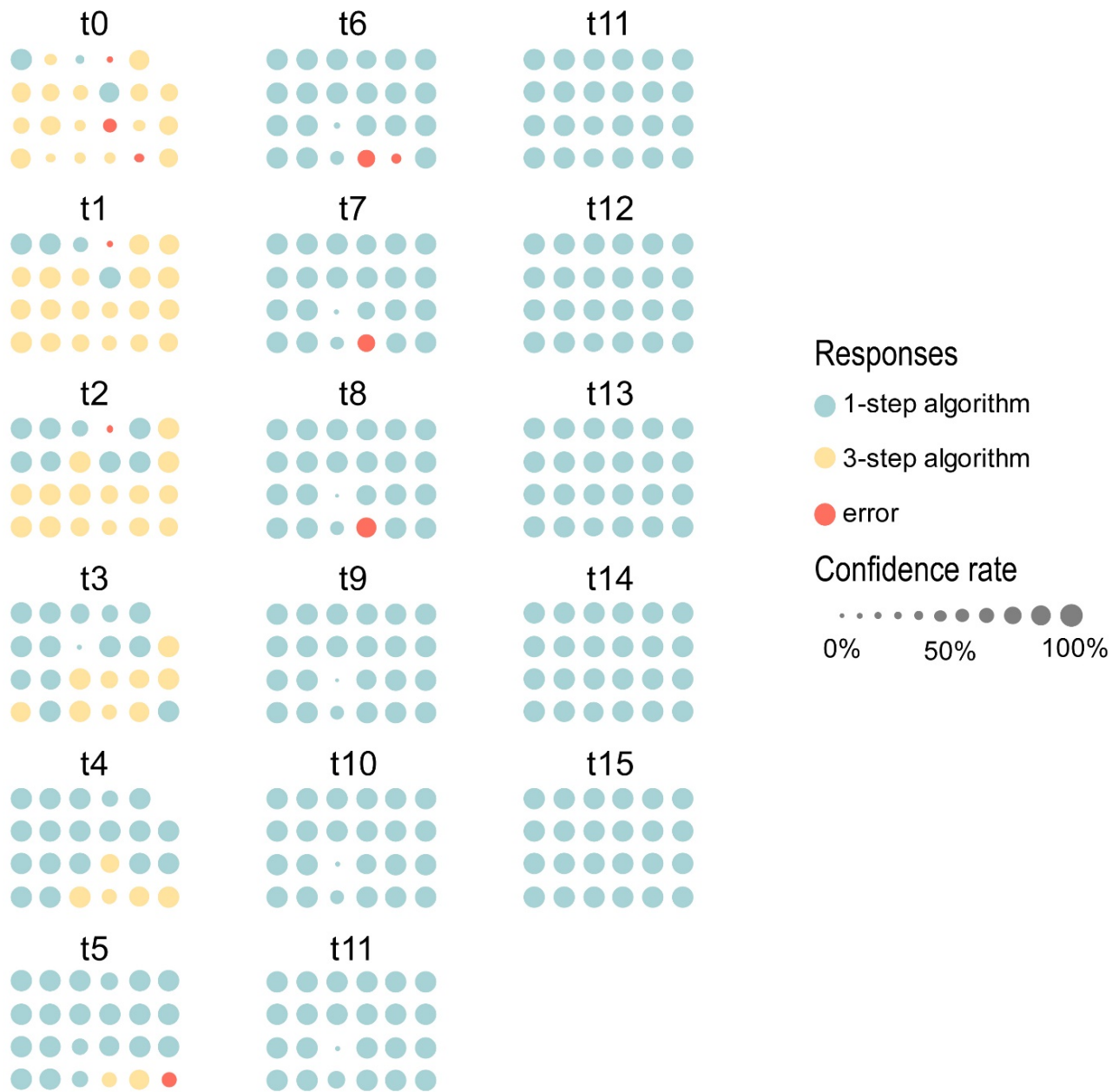


Fig. 2 Students' responses at each time stamp for the 1st cardinal problem. Each circle represents a student in the classroom. The color of the circle indicates the strategy used, the circle size indicates confidence rate.

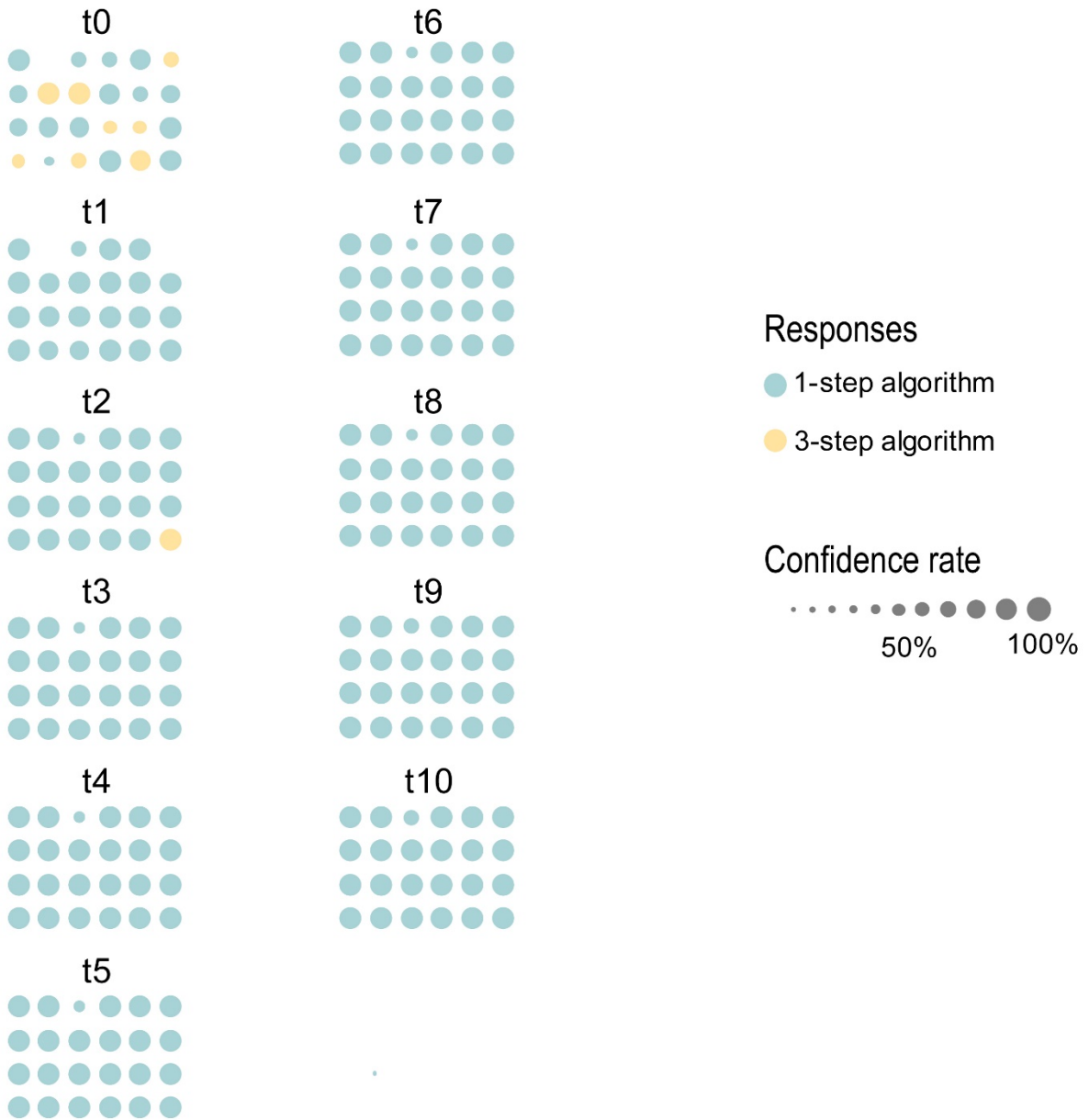


Fig. 3 Students' responses at each time stamp for the 2nd cardinal problem. Each circle represents a student in the classroom. The color of the circle indicates the strategy used, the circle size indicates confidence rate.

Conclusion

If we had to highlight one take-home message from this thesis, it would probably be that humans reason differently depending on what they reason about, even in domains where abstraction is trained and valued. In our case, this idea took the form of participants using one algorithm to solve a smurf problem, and another to solve an elevator problem. But the more general notion behind it, the view that our domain-independent knowledge shapes the representations underlying our reasoning and subsequent behavior, is one that has a much larger scope. Similarly to how non-mathematical knowledge interfered with mathematical reasoning in our experiments, domain-independent knowledge may meddle with human reasoning in a wide range of domains and activities. This idea raises crucial issues for education, as it emphasizes the notion that learning and teaching do not happen in a vacuum, but are always grounded in a real-world setting. This inter-dependency of context-specific and context-independent knowledge may be a source of confusion at times, but it may also be a driving force helping individuals comprehend complex abstract ideas with little effort, through semantic congruence. Overall, we believe that it is time to let go of the idea that humans should strive above all to act as purely rational beings, guided by reason alone. The pervasive interferences between our intuitions and our expertise, our beliefs and our judgment, our experience and our perception, our skills and our wisdom, our memories and our comprehension, all shape our cognitive processes, and define us as human beings.

References

- Baruk, S. (1985). *L'âge du capitaine. De l'erreur en mathématiques*. Berlin: Springer.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. California University Press.
- Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43(1), 1-22.
- Caldwell, L. (1995). *Contextual considerations in the solution of children's multiplication and division word problems*. Unpublished Master's thesis, Queen's University, Belfast, Northern Ireland.
- Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language learning*, 58, 63-79.
- Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, I., ... Gil, D. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. In: K. Forbus, D. Gentner, T. Regier, (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates, Hillsdale, NJ, 2004, pp. 575-580.
- Casasanto, D., Fotakopoulou, O., & Boroditsky, L. (2010). Space and time in the child's mind: Evidence for a cross-dimensional asymmetry. *Cognitive science*, 34(3), 387-405.
- Chomsky, N. (1975). *The logical structure of linguistic theory*. Berlin: Springer.
- Claidière, N., Trouche, E., & Mercier, H. (2017). Argumentation and the diffusion of counter-intuitive beliefs. *Journal of Experimental Psychology: General*, 146(7), 1052.
- Csikós, C. (2003, August). How many buses are needed? Hungarian students' achievement on 'problematic' word problems. In *10th European Conference for Research on Learning and Instruction, Padova, Italy*.
- DeFranco, T. C., & Curcio, F. R. (1997). A division problem with a remainder embedded across two contexts: Children's solutions in restrictive versus real-world settings. *Focus on Learning Problems in Mathematics*, 19(2), 58-72.
- Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *science*, 320(5880), 1217-1220.

- Eagleman, D. M. (2001). Visual illusions and neurobiology. *Nature Reviews Neuroscience*, 2(12), 920.
- Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard university press.
- Gamo, S., Nogry, S., & Sander, E. (2014). Apprendre à résoudre des problèmes en favorisant la construction d'une représentation alternative chez des élèves scolarisés en éducation prioritaire. *Psychologie française*, 59(3), 215-229.
- 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.
- Greer, B. (1993). The modelling perspective on wor(l)d problems. *Journal of Mathematical Behavior*, 12, 239-250
- Guthormsen, A. M., Fisher, K. J., Bassok, M., Osterhout, L., DeWolf, M., & Holyoak, K. J. (2016). Conceptual integration of arithmetic operations with real-world knowledge: Evidence from event-related potentials. *Cognitive science*, 40(3), 723-757.
- Hidalgo, M. C. (1997). *L'activation des connaissances à propos du monde réel dans la résolution de problèmes verbaux en arithmétique*. Unpublished doctoral dissertation, Université Laval, Québec, Canada.
- Kounios, J., & Beeman, M. (2009). The Aha! moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science*, 18(4), 210-216.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*, 4(12), 463-470.
- Mercier, H. (2016). The argumentative theory: Predictions and empirical evidence. *Trends in Cognitive Sciences*, 20(9), 689-700.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57.
- Oh, K. J. (2003). Manner and Path in motion event descriptions in English and Korean. In *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 580-590). Boston, MA: Cascadilla Press.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 786.

- Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, 84(2), 189-219.
- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306(5695), 499-503.
- Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: Harper.
- Pullum, G. K. (1991). *The great Eskimo vocabulary hoax and other irreverent essays on the study of language*. Chicago: University of Chicago Press.
- Renkl, A. (1999). *The gap between school and everyday knowledge in mathematics*. Paper presented at the Eighth European Conference for Research on Learning and Instruction, Göteborg, Sweden, August 1999.
- Reusser, K., & Stebler, R. (1997). Every word problem has a solution: The suspension of reality and sense-making in the culture of school mathematics. *Learning and Instruction*, 7, 309-328.
- Reys, R., Reys, B., Emanuelsson, G., Johansson, B., McIntosh, A., & Yang, D. C. (1999). Assessing number sense of students in Australia, Sweden, Taiwan, and the United States. *School Science and Mathematics*, 99(2), 61-70.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9(1), 76-80.
- Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge: Cambridge University Press.
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9(1), 59-71.
- 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.
- Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modelling of school arithmetic word problems. *Learning and Instruction*, 4, 273-294.

- Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Design and evaluation of a learning environment for mathematical modelling and problem solving in upper elementary school children. *Mathematical Thinking and Learning, 1*, 195–229.
- Verschaffel, L., Van Dooren, W., Greer, B., & Mukhopadhyay, S. (2010). Reconceptualising word problems as exercises in mathematical modelling. *Journal für Mathematik-Didaktik, 31*(1), 9-29.
- Whorf, B. L. (1939). The relation of habitual thought and behavior to language. In J. B. Carroll (Ed.), *Language, thought and reality: Selected writings of Benjamin Lee Whorf* (pp. 134–159). Cambridge, MA: MIT Press.
- Xin, Z., Lin, C., Zhang, L., & Yan, R. (2007). The performance of Chinese primary school students on realistic arithmetic word problems. *Educational Psychology in Practice, 23*, 145–159.
- Yoshida, H., Verschaffel, L., & De Corte, E. (1997). Realistic considerations in solving problematic word problems: Do Japanese and Belgian children have the same difficulties? *Learning and Instruction, 7*, 33.