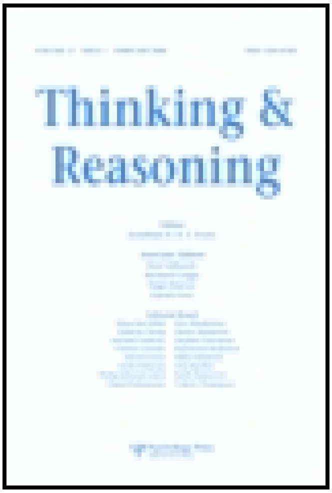


Thinking & Reasoning
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ptar20

Contraries as an effective strategy in geometrical problem solving

Erika Branchini[a], Roberto Burro[a], Ivana Bianchi[b] & Ugo Savardi[a]

[a] Department of Philosophy, Education and Psychology, University of Verona, Verona, Italy

[b] Department of Humanities, University of Macerata, Macerata, Italy

Published online: 03 Jan 2015.

To cite this article: Erika Branchini, Roberto Burro, Ivana Bianchi & Ugo Savardi (2015): Contraries as an effective strategy in geometrical problem solving, Thinking & Reasoning, DOI: 10.1080/13546783.2014.994035

To link to this article: http://dx.doi.org/10.1080/13546783.2014.994035

Thinking & Reasoning, 2014
http://dx.doi.org/10.1080/13546783.2014.994035

Contraries as an effective strategy in geometrical problem solving

Erika Branchini[1], Roberto Burro[1], Ivana Bianchi[2], and Ugo Savardi[1]

[1]Department of Philosophy, Education and Psychology, University of Verona, Verona, Italy
[2]Department of Humanities, University of Macerata, Macerata, Italy

A focused review of the literature on reasoning suggests that mechanisms based upon contraries are of fundamental importance in various abilities. At the same time, the importance of contraries in the human perceptual experience of space has been recently demonstrated in experimental studies. Solving geometry problems represents an interesting case as both reasoning abilities and the manipulation of perceptual−figural aspects are involved.

In this study we focus on perceptual changes in geometrical problem solving processes in order to understand whether a mental manipulation in terms of opposites might help. Four conditions were studied, two of which concerned the search for contraries as an implicit or explicit strategy.

Results demonstrated that contraries, when used explicitly in solution processes, constitute an effective heuristic: The number of correct solutions increased, less time was needed to find a solution and participants were oriented towards the use of perception-based solutions—not only were perceptual solutions more frequent, but also, more specifically, the number of correct perceptual solutions increased. These last results concerning perception-based solutions were found both when participants were advised about the usefulness of the strategy and when they were not advised. Differences concerning which aspects of a problem were focused on during the solution process also emerged.

Keywords: Contraries in perception and reasoning; Geometrical problem solving; Contrast class; Opposites.

Contraries have been widely investigated in studies by contemporary researchers in psycholinguistics and cognitive linguistics. These studies

Correspondence should be addressed to Ivana Bianchi, Department of Humanities, University of Macerata (Italy), via Garibaldi 20, 62100 Macerata, Italy. E-mail: ivana.bianchi@unimc.it

We would like to thank two anonymous reviewers and the Editor, Linden Ball, for their valuable comments.

No potential conflict of interest was reported by the authors.

showed that antonyms—intended as a cover term for all types of oppositeness (Crystall, 1985)—are a typical structure common to all natural languages. Up/down, big/small, good/bad, old/young, etc. are examples of everyday antonymic words.

Researchers in psycholinguistics and cognitive linguistics have always emphasised the importance and primacy of opposites in human semantics and conceptual spaces (Croft & Cruse, 2004; Jones, 2002; Jones, Murphy, Paradis, & Willners, 2012; Murphy, 2003; Paradis, Hudson, & Magnusson, 2013; Paradis & Willners, 2011). And now other recent studies have pointed out the importance of opposites in perception and reasoning. For instance, it has been shown that:

(1) opposites constitute a primal perceptual relation, different from diversity and similarity, and they are particularly prominent in spatial perception. This has emerged from both visual and motor studies (Bianchi, Burro, Torquati, & Savardi, 2013; Bianchi & Savardi, 2008; Bianchi, Savardi, & Burro, 2011; Bianchi, Savardi, Burro, & Martelli, 2014; Bianchi, Savardi, & Kubovy, 2011; Savardi & Bianchi, 2009). The hypothesis of a pre-linguistic foundation to the importance of opposites in cognition is also supported in studies regarding cognition in infants (Casasola, 2008; Casasola, Cohen, & Chiarello, 2003);
(2) opposites positively affect reasoning in hypothesis testing. The standard paradigm employed to study performance in hypothesis testing is Wason's rule discovery task (Wason, 1960). It has been shown that reasoning in terms of contrast class cues is an effective strategy to help overcome the confirmatory bias which typically characterises the process of searching for a solution (Gale & Ball, 2006, 2009, 2012).

Using this as a starting point, in this paper we put forward and test the hypothesis that reasoning in terms of contraries may help in problem solving beyond the conditions considered in typical studies investigating the efficacy of contrast class cues. Before presenting our study, we will briefly review the literature which has demonstrated that problem solving processes are sensitive to visual or visually based manipulation of a problem situation. This is relevant to the present paper since our study concerns geometrical problem solving, i.e., problems where figurative aspects (either under observation or imagined) are essential.

THE STATE OF THE ART REGARDING THE RECOGNITION OF THE ROLE OF CONTRARIES IN HYPOTHESIS TESTING

In the classic task devised by Wason (1960), participants are invited to discover the rule, known only to the experimenter, that governs the production of triple numbers. Starting from the sequence 2-4-6, considered as a seed triple, participants are asked to generate further triple numbers which they consider useful in order to discover the rule (the to-be-discovered rule is "any ascending sequence"). The experimenter confirms each time whether the proposed triple number conforms or not to the target rule. Participants are only allowed one attempt and thus announce that they have discovered the rule only once they are confident. Despite the apparent simplicity of the task, only a small percentage of participants are successful and some remain unable to discover the rule even after repeated attempts (Farris & Revlin, 1989b; Wason, 1960). Wason's original view (e.g., Wason & Johnson-Laird, 1972) was that people succumb to a confirmation bias on the task, which arises from the use of a verification strategy whereby they propose triples that potentially conform to the hypothesis they have in mind, rather than trying to falsify it. However, more recent evidence (e.g., Klayman & Ha, 1987, 1989) has recast behaviour on the task as reflecting the use of a "positive test strategy" whereby people generate positive tests of their hypotheses, which inadvertently leads them into loops of confirmation because of the peculiar properties of the task rather than because they lack the motivation or ability to falsify (see Evans, 2014). This re-framing of the evidence perhaps explains why researchers who have implemented various falsification interventions to improve performance have generally been unsuccessful (Gorman, Stafford, & Gorman, 1987; Kareev, Halberstadt, & Shafir, 1993; Tweney et al., 1980).

Improvements were obtained with a variation of Wason's task (Tweney et al., 1980), with participants being asked to discover *two* rules, one (called DAX) was "any ascending sequence of numbers", the other (called MED) "all other sequences" (i.e., any non-ascending sequence). A number of studies demonstrated that this dual task facilitates the discovery of the rules (Gale & Ball, 2006; Gorman et al., 1987; Tukey, 1986; Vallée-Tourangeau, Austin, & Rankin, 1995; Wharton, Cheng, & Wickens, 1993). However, various theoretical explanations of why this happened turned out to be unsatisfactory—for example, Evans' proposal to positively label "does not fit" feedback (Evans, 1989), the goal complementary theory (Wharton et al., 1993) and the triple-heterogeneity theory (Vallée-Tourangeau et al., 1995). The key factor seems rather to be the use of contrast class cues, as shown by studies demonstrating that a facilitating mechanism emerged when cues based upon the oppositional nature of DAX and MED were provided (Gale & Ball, 2003, 2006, 2009, 2012).

The definition of a "contrast class cue" was introduced by Oaksford and Stenning (1992; Oaksford, 2002), who clarified that it is a *psychological* and not logical concept in that the idea of contrast class does not refer to the strict complement of a set but to its most likely or relevant members. For instance, if we are told that "John was not drinking a coffee", we are inclined to hypothesise (driven by what the authors call "perceived relevance") that John was drinking another hot beverage, e.g., tea, hot chocolate or Ovaltine, rather than one of the multitude of other drinks that would fall into the logical complement, e.g., whisky, water, cola, etc. By applying this definition to the problem involving triple numbers, Gale and Ball (2012) demonstrated that the 6-4-2 triple was the MED example that fit in best with the peculiarities of the useful contrast class and in fact it greatly facilitated the discovery of the contrasting DAX (ascending) rule. Conversely, 4-4-4 and 9-8-1 were triples which were logically coherent but psychologically misleading and they thus inhibited the discovery of the DAX (ascending) rule. The 4-4-4 triple was useless since it biased participants towards an irrelevant dimension (i.e., identical vs. different numbers) and hid the importance of the ascending vs. descending dimension. The 9-8-1 triple did not facilitate the search for the DAX rule since it contrasted with the 2-4-6 triple in several dimensions in addition to the relevant descending–ascending dimension. These irrelevant dimensions were, for example: Odd and even numbers vs. only even numbers, unequal vs. equal intervals and a middle number which is the arithmetic mean of the other two numbers vs. one which is not. This profusion of contrasting elements hinders the emergence of the critical contrast.

GOING BEYOND WASON'S TASK: THE IMPORTANCE OF THE VISUAL STRUCTURE OF THE PROBLEM IN GEOMETRICAL UNDERSTANDING AND PROBLEM SOLVING

Wason's task is for sure a typical example of a problem solving condition. Problem solving has been defined as a particular thought process activated by human beings when they have to cope with situations characterised by a goal to be reached and by the search for the means to pursue it (Anolli, Antonietti, Crisafulli, & Cantoia, 2001; Johnson-Laird, 1993; Newell & Simon, 1972).

Gestalt psychologists (Duncker, 1926,1935, 1945; Köhler, 1925; Maier, 1930, 1931a, 1931b, 1945; Wertheimer, 1919/1945) provided a phenomenological description of what happens when people face a problematic situation (see Nerney, 1979). The distinction between productive thinking (i.e., what characterises all intelligent acts that create something new) and reproductive thinking (i.e., the application of a previously known chain of associations, learnt in the past and reinforced by habits) was central to their analysis. Their primary contribution was to demonstrate the importance of paying attention to the "phenomenal structure" of a problem (Luchins & Luchins, 1970a, 1970b, 1970c; Wertheimer, 1919/1945), i.e., how the structure of a problem is spontaneously perceived by problem solvers, what its primary and secondary elements appear to be and what the relationships between these are. This initial phenomenal structure may be a kind of "trap" which prevents the problem solver from seeing the solution (Duncker, 1926, 1935, 1945; Harrower, 1932; Köhler, 1969)[1] but it also potentially shows the gaps or "trouble zones" to be healed, i.e., it may also suggest the directions to follow in order to solve the problem (Wertheimer, 1919/1945). Note that productive thinking was not thought of by gestalt psychologists as a cognitive process involving higher order cognitive abilities but rather as a case of perceptual processing: Understanding the phenomenal structure meant *seeing* the phenomenal organisation of the problematic elements and the necessary procedure to solve them. This may mean, for example, reassessing the part–whole organisation of a problem involving the division of elements that were grouped together in the original spontaneous phenomenal organisation and, vice versa, the unifying of components that initially appeared to be separated.

The role of the "visual structure" of a problem in facilitating or impeding problem solving was reconfirmed in three more recent streams of studies. First, its importance emerged in studies which demonstrated the importance of visualisation and the ability to manipulate mental images when learning geometry (Duval, 2006; Gorgorio, 1998; Gutiérrez, 1996; Jones & Bills, 1998). Dividing the figure into new parts or sub-parts (mereological transformations), making the figure wider or narrower as if it were being viewed through a lens and manipulating different planes of spatial representation (thus transforming the orientation of the figure with respect to the observer) are three visual operations which have turned out to be critical in terms of helping people understand geometrical problems and demonstrations (Duval, 2006; Gutiérrez, 1996; Gutiérrez, Pegg, & Lawrie, 2004; Herbst, 2006; Unal, Jakubowski, & Corey, 2009). Visualisation also helps in the discovery of mathematical truths (Giaquinto, 2007); for example, the general proposition regarding squares in Euclidean space that "any square C has twice the area of the square whose vertices are midpoints of C's side" is easier to believe if one simply visualises it (imagine a square with a horizontal base: Each of its four sides has a midpoint and there is a second square inside the first square whose corner points coincide with the four midpoints of the first square; this second inner square is tilted diagonally so it seems to stand on one of its corners; visualisation makes it easier to see that the original square equals twice the area of the inner square because it is composed of the tilted square plus four other corner triangles, each with one side of the inner square as its base; the corner triangles can be "folded over" the sides

[1] The biases related to remaining attached to the usual functions of everyday objects rather than seeing new possible uses of them (Duncker, 1926, 1935, 1945) and also imputing a procedure already repeated in a series of similar tasks to subsequent problems (known as the Einstellung effect by Maier, 1930, 1931a, 1931b, 1945; Luchins, 1942, 1946; Luchins & Luchins, 1950) are two variants of this.

of the inner square thus covering the area of the inner square exactly, without any gap or overlap).

Second, studies on insights in problem solving have demonstrated that before the insight arrives, problem solvers usually encounter many hindrances which could potentially cause an *impasse* (i.e., a state of mind in which one does not know what to do next) and which depend on the initial representation of the problem. Problem solvers frequently hark back to past experiences and as a result perceptually organise the elements of a problem according to familiar perceptual patterns. However, this means that they apply familiar categories, concepts, rules or schemas and these somehow "constrain" the options which they initially considered and thus limit the area within which they search for a solution. This initial representation activates some elements of knowledge but not others and thus inhibits the discovery of further components which are not associated with the initial representation but may be potentially of interest. The impasse can be broken by a representational change (Knoblich, Ohlsson, & Raney, 2001; Ohlsson, 1992; Öllinger, Jones, & Knoblich, 2008; Wu, Knoblich, & Luo, 2013) and this is a cognitive process which implies visually based processes (Luo, Niki, & Knoblich, 2006).

Third, evidence of the influence of visual experiences in problem solving also emerged in a set of studies demonstrating that the use of visual images or diagrams facilitates the understanding of a situation and the discovery of a solution in classic problem solving tasks, such as for example Wason's task (Vallee-Tourangeau & Pyton, 2008)[2] or Maier's problem[3] (Antonietti, 2001; Antonietti, Angelini, & Cerana, 1995).

THE HYPOTHESIS UNDERLYING OUR STUDY

From the above-mentioned literature a series of considerations can be drawn. (1) In the problem solving process, problem solvers often need to go beyond the initial representation of the problem that takes shape in their mind or the initial solution they come up with. (2) This initial representation of the problem situation is also defined by visual aspects (mostly spatial) which are either actually under observation or mentally visualised or both. This is generally true but is particularly prominent when geometrical problem solving is concerned. (3) We know from studies on direct experiences of space and phenomenal representations of space that spatial experience is

[2] Representing numbers as points on a line is an example of representing abstract concepts in concrete terms (for a theory of embodied mathematics see Lakoff & Nunez, 2000).

[3] This problem (1930) asks people to make four equilateral triangles out of six matches, with each triangle having one whole match for each side: Visualizing the matches arranged as a three-dimensional tetrahedron (each face corresponding to a triangle) significantly facilitates the task.

inherently oppositional (Bianchi et al., 2013; Bianchi, Savardi, & Kubovy, 2011; Savardi & Bianchi, 2009): *x* is perceived as either *near* or *far* from *y* (or *neither near nor far*, which is in any case an intermediate state in a dimension defined by two opposite poles), or *x* can be either *inside* or *outside*, *open* or *closed*, *above* or *below*, *big* or *small*, *ascending* or *descending*, etc. This oppositional schema gives form to the way in which we conceive space to be, both in general (e.g., Gardenfors, 2000, 2014) and with regard to specific phenomena (such as our understanding of mirror reflections for example; see Savardi, Bianchi, & Bertamini, 2010).

Manipulating the spatial characteristics of a problem in terms of contraries should thus on the one hand be *easy* and on the other hand should offer an *effective* strategy to go beyond the initial representation of the problem; it in fact suggests a way to *systematically* operate on the situation to explore transformations leading to possible solutions. Reasoning in terms of contraries prompts problem solvers to focus, at different moments, on one or the other of the aspects of a problem and to wonder whether the state of things they have in mind is necessarily so or if the solution might require transforming this state into its opposite. This may be in terms of extension (e.g., wide−narrow, high−low, long−short, etc.), position (e.g., above−below, inside−outside, to the left−to the right, beginning−end, etc.), orientation (e.g., vertical−horizontal, convergent−divergent, ascending−descending, etc.) or shape (open−closed, symmetrical−asymmetrical, straight−bent, etc.). For example, is the fact that a figure is positioned horizontally (or the fact that the segments constituting its sides are unified to form a closed shape etc.) a necessary constraint? Or rather, might the solution require the figure to be positioned vertically or the segments forming the figure's sides to be separate and no longer form a closed contour?

Note that we are not implying that people will immediately realise *which* transformation will lead to the solution. We expect this strategy to be useful to help people work productively on the information available in order to eliminate constraints. Moreover this is a strategy that does not require expert knowledge and in this sense it is potentially available to any problem solver (which means that people might take advantage of it not only individually but also when cooperating with other non-experts).

To exemplify how contraries might help problem solvers to restructure a problem, let us consider Wertheimer's parallelogram problem (Wertheimer, 1945) and Kanizsa's square problem (Kanizsa, 1973).[4] In the parallelogram problem the task is to calculate the area of a parallelogram; the solution consists of transforming the parallelogram into a rectangle (see the top row in Figure 1). This global change implies the use of contraries because it

[4] Some of us (Branchini, Burro, & Savardi, 2009) have analysed 10 classic problems used in research on problem solving to show that the solution processes often require the transformation of some of the initial properties of a problem into their contraries.

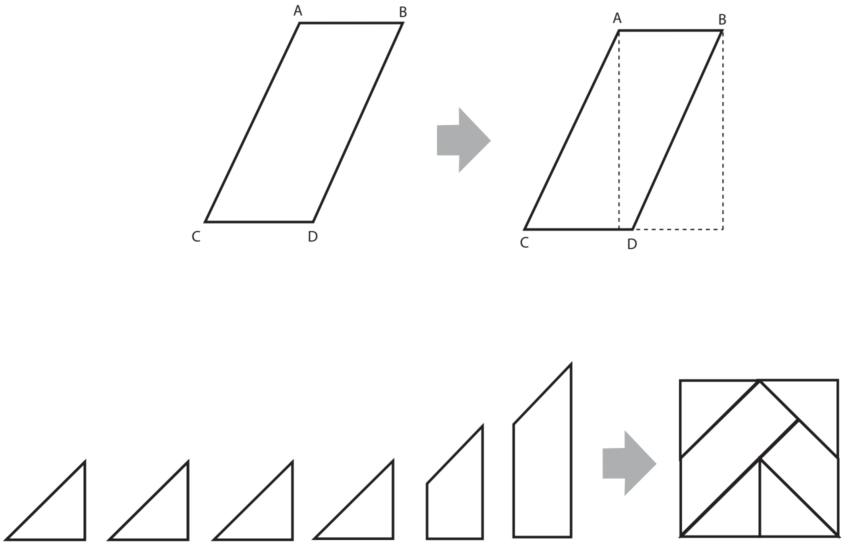


Figure 1. Initial configurations and solutions of Wertheimer's parallelogram problem (first row) and Kanizsa's square problem (second row).

requires the transformation of two salient global characteristics in the initial figure—i.e., it is visually "unstable" and leans horizontally—into the opposite of those characteristics to produce a stable, upright figure. This is achieved by drawing two perpendicular lines down from the A and B vertexes to form two identical triangles. This in turn leads to other variations involving contraries. One triangle is on the left, inside the parallelogram and the other is on the right, outside the parallelogram. By moving the left hand triangle (which is outside the rectangular shape and thus "in excess") to fill in the space on the right (which appears to be "missing" from the rectangle), one obtains a rectangle that is perfectly equivalent to the parallelogram. The area can then be easily calculated by multiplying the base by the height.

Similar considerations can be made regarding Kanizsa's square problem. The task is to build a square by putting together six figures (four right-angled isosceles triangles and two right-angled trapezoids with bases of different sizes but of equal height—see the bottom row in Figure 1). Not only does the problem itself present us with an initial element of contrariety (i.e., the task is to transform *many* separate shapes into *a single* unified shape) but the solution also implies the manipulation of various spatial contraries in order to change the initial orientation and position of the separated shapes. The two trapezoids, which are vertical when separated, must be positioned obliquely in the new figure. We see that the horizontal and vertical sides of the trapezoids are not in fact used to build the horizontal and vertical sides of the new square but, conversely, are placed obliquely inside the new square to be lined up with the sides of the triangles. The oblique sides of the trapezoids in this way form part of the four sides of the new square. In addition, the four triangles which are identically oriented in the initial configuration have various different orientations in the new configuration: The two triangles at the top are positioned symmetrically but arranged in such a way that they only touch each other at the vertex and the two right-angle vertexes are external, forming two of the vertexes of the new square; in contrast, the two other triangles at the bottom are again symmetrically positioned but adjacent, with an entire side in common, and the two right-angle vertexes are internal in the new set-up.

A pilot study with children (briefly presented in Branchini, Burro, & Savardi, 2009) provided the first evidence in support of the positive effects of encouraging problem solvers to look for contraries. In the present study we tested this hypothesis with adults, using a more complete experimental design and a wider sample of participants.

THE STUDY

The aim of the study was to see what happened when we asked adult participants to look for contraries when trying to solve six geometrical problems. In one condition we explicitly stated that this might help them with the solution process, in another condition we left this unsaid. The reason for this was the need to understand whether carrying out a preliminary exercise which activated the exploration of the situation in terms of oppositional dimensions positively impacted on the solution process or whether this was effective only when the participants were prompted to apply this strategy in order to reach the goal. These two conditions were compared with two other conditions: A control condition where no indications were provided and a further condition where participants were prompted to recall their knowledge of geometry.

For all four conditions we analysed: The type of solution found, the time needed to reach a solution, and the type of processing that was carried out on the data related to the problem. This latter aspect involved observing which aspects of a problem the participants focused on while searching for a solution. This was the most difficult information to capture since it required us to somehow look into the participants' minds while they were processing the data. In order to address this, we combined two different methods: The inter-observational method and the "thinking aloud coding scheme" method. The former was introduced by experimental phenomenologists of perception to obtain precise and refined descriptions of an event under observation (Bianchi, Savardi, & Kubovy, 2011; Bozzi, 1978; Bozzi & Martinuzzi, 1989; Kubovy, 2002). It requires participants to take part in an experiment in small groups of three to five members and to share their thoughts by giving verbal descriptions of how they see the situation. This activates a process of cross-testing of each individual detail of the descriptions given in order to check whether they fit in with what all the other observers have seen. The second method, the "thinking aloud coding scheme", comes from work domain analysis (Hoffman & Lintern, 2006; Rasmussen, 1985; Rasmussen, Petersen, & Goodstein, 1994; Rasmussen, Petersen, & Schmidt, 1990). It consists of an "abstraction−decomposition" matrix, used to analyse the representations and processes that characterise expert knowledge. Expert knowledge can be accessed using different methods (e.g., structured and unstructured interviews, ethnographic studies of patterns of communication emerging in workplaces, performance measures, reaction times in contrived tasks or think-aloud problem solving tasks, etc.). The matrix explores implicit knowledge using a list of categories whose contents vary depending on the specific expertise being studied. This expertise is closely examined in terms of how abstract or concrete aspects are processed (levels of abstraction) and whether it operates on a local or a global level (levels of decomposition).

In our study, the two methodologies were jointly applied: Participants took part in the study in small groups of three people, and were invited to solve the task collectively by thinking aloud and sharing their thoughts. This allowed us to obtain information regarding the search for a solution since this was "an in progress process" (see "Procedure" section for more details). The plausibility of this choice was also supported by results from previous studies which showed that either there are no overshadowing effects of verbalisation on insights in problem solving (Ball & Stevens, 2009; Fleck & Weisberg, 2013; Gilhooly, Fioratou, & Henretty, 2010) or that if these arise during a short period of time they will dissipate eventually over longer times (Ball, Marsh, Litchfield, Cook, & Booth, 2014)—and in our study participants were given unlimited time.

METHOD

Participants

240 students (96 undergraduate students and 144 final year high school students) divided into 80 inter-observational groups, each with three members.

Procedure

Participants took part in the study in groups in a meeting room at the University of Verona (the undergraduate students) or an empty classroom in a secondary school in the same region of Italy (the high school students). The groups were provided with a seven-page booklet with instructions on the first page. These were followed by the six problems (each one on a separate page), in random order. The instructions varied in the four conditions studied.

In one condition (Advised use of Contraries: AC) participants were instructed to look for contraries and were advised that this strategy would help the solution process. "Before searching for the solution to the problem, we would like you to perform a preliminary task: List all the spatial contrary features (e.g., at the top–at the bottom) which are explicitly present in the problem or are "implicit". By "implicit" we mean that only one of the two poles is present, either because it is specifically mentioned (e.g., "at the top" is referred to but not "at the bottom") or because the objects referred to are characterised by it (e.g., if there is a star-shaped figure, this implies angularity, spikiness or broken contours). When only one pole is present, try to identify its contrary. Make an exhaustive list of these contraries, write them down. Once you have finished, you can move to the second phase and look for a solution. Do the initial phase carefully since it will help you to find a solution in the second phase".

In a second condition (non-Advised use of Contraries: nAC) participants were instructed to look for contraries but were not advised that this strategy would help the solution process. The instructions were the same as above, except for the last sentence which was omitted.

In a third condition (Advised use of prior geometrical Knowledge: AK) participants were instructed to use previous geometrical knowledge and were advised that this strategy would help the solution process. ("In your search for a solution, use any knowledge, competence, rules and concepts that you have learned in the past and which you consider are needed to solve the problem".)

In a fourth condition (no strategy: n) participants were given no instructions regarding the use of a strategy and no advice that a strategy would help.

In all four instruction conditions, the participants were invited at the beginning of the session to solve the task collectively by talking aloud and sharing their thoughts. It was made clear that they were allowed to use paper, pens, pencils and rulers during the session and that they could do the problems presented in the booklet in any order. Only one booklet per group was provided in order to stimulate them to work together. It was also specified that there were no time limits. Each experimental session took place in the presence of an experimenter and was video recorded, with their permission for this having been requested previously.

Material

A booklet consisting of seven A4 sheets of paper was given to each group. The instructions were printed on the first page and the details of the six spatial-geometrical problems (and the corresponding figures where appropriate) were printed on the following six pages. The problems (reported in the Appendix) were Wertheimer's parallelogram (Wertheimer, 1945), Maier's nine dots (Maier, 1930), Harrower's duck (Harrower, 1932), Wertheimer's altar window (Wertheimer, 1945), Kanizsa's square (Kanizsa, 1973) and Köhler's circumference (Köhler, 1969).

RESULTS

Type of solution

We first of all studied whether there were any differences between the four different instructions in terms of the type of solution that participants came up with. We distinguished five types of solutions and classified responses accordingly: *Correct perceptual solutions* (cPS), i.e., correct solutions based on perceptual spatial operations, without the application of previously learned rules; *incorrect perceptual solutions* (iPS), i.e., incorrect solutions based on perceptual spatial operations, without the application of previously learned rules; *correct prior knowledge solutions* (cKS), i.e., correct solutions resulting from the application of previously learned knowledge or previously known rules; *incorrect prior knowledge solutions* (iKS), i.e., incorrect solutions based on the application of previously learned knowledge or previously known rules and *no solution* (nS) referring to the cases where participants did not succeed in finding a solution (either correct or incorrect). Two independent judges did the codification and there was a remarkable agreement between their evaluations (Cohen's K index = 0. 86).

The percentage of groups that did not find a solution was less than 1% (0.8%). These responses were eliminated from any further analyses which thus focused exclusively on the types of solutions found. The frequency of the various solution types is shown in Table 1. Two initial considerations can be put forward based on the totals. First, perceptual solutions (independently of whether correct or incorrect) represented the most frequent type of solution: They were 55% of the final solutions offered by participants vs. 45% of solutions based on previous geometrical knowledge (PS = 264 vs. KS = 212; McNemar chi-square = 5.46, $p < .02$). This might not be surprising given the type of problem presented. A second more meaningful finding

TABLE 1
Type of solution (correct prior knowledge solution, cKS; correct perceptual solution, cPS; incorrect prior knowledge solution, iKS; and incorrect perceptual solution, iPS) reached by participants in the four instruction conditions: Advised use of Contraries (AC), Advised use of prior geometrical Knowledge (AK), non-Advised use of Contraries (nAC) and no strategy (n)

			cKS	*cPS*	*iKS*	*iPS*	*Total*
Instructions	AC	Counts	25	70	16	7	118
		% within AC	21.2%	59.3%	13.6%	5.9%	100.0%
	AK	Counts	39	46	26	8	119
		% within AK	32.8%	38.7%	21.8%	6.7%	100.0%
	nAC	Counts	22	62	23	12	119
		% within nAC	18.5%	52.1%	19.3%	10.1%	100.0%
	n	Counts	30	50	31	9	120
		% within n	25.0%	41.7%	25.8%	7.5%	100.0%
Total	Counts		116	228	96	36	476
	%		24.3%	47.9%	20.2%	7.6%	100.0%

concerns the proportion of correct and incorrect solutions within the two types of responses: Perceptual solutions were more frequently correct than incorrect (cPS = 228, iPS = 36; McNemar chi-square = 138.19, $p < .0001$) while solutions based on prior knowledge turned out to be as frequently incorrect as correct (cKS = 116; iKS = 96; McNemar chi-square = 1.70; $p = 0.19$). This suggests that not only are PS often chosen, but also that this choice is often appropriate.

In the following analyses, data were studied using generalised mixed effects models (GLMMs – R package "lme4", version 1.1-7; Bates, Maechler, Bolker, & Walker, 2014a, 2014b). In fact our experimental design included both random effects (Groups and Problems) and fixed effects (Instructions) and various types of dependent variables (binomial and scale values).

Correct vs. incorrect.. The first question we were interested in was whether correct responses were more frequent in any of the four instruction conditions. In order to test this, we recoded responses as correct (C) or incorrect (I)—independently of whether they were perceptual solutions or prior knowledge solutions—and carried out the following tests on this new binary dependent variable using a logit link function.

To begin with, we studied the effects of our two random effects (Groups and Problems) by testing the difference between the more complex model (which includes both random intercepts and random slopes, i.e., both intercepts and slopes are allowed to vary across Groups and across Problems) and the simpler model which includes only random intercepts. In addition, Instruction was inserted in both models as a fixed effect. No significant differences emerged between the two models (chi-square = 0.325, df = 18, p = 1.000). Akaike's information criterion (AIC) was used to measure the relative quality of these models for our set of data—AIC offers a relative estimate of the information lost when a given model is used to represent the process that generates the data. Since AIC was smaller in the simpler model (AIC Instruction + random intercepts$_{(\text{Groups, Problems})}$ = 410.31; AIC Instruction + random intercepts and slopes$_{(\text{Groups, Problems})}$ = 445.99), we used it to model the random effects in the next analysis where the fixed effect (Instruction) was eliminated. Instructions turned out to be a relevant factor in modelling data variability. Deleting it led to a significantly worse fit of the model to the data (chi = 7.6832, df = 3, p = 0.05). The AIC confirmed that in terms of fit this model is worse than that which includes Instructions (AIC random intercepts = 411.40).

In Figure 2 the plot of the effect of Instructions is shown. Let us bear in mind that in a model like the one used here (which works on a logit transformation of a binary variable, in our case correct vs. incorrect solutions), an equal proportion of the two categories of the binary variable (i.e., the number of correct solutions = the number of incorrect solutions) corresponds to a logit value of 0.5, whereas logit values greater than 0.5 indicate that

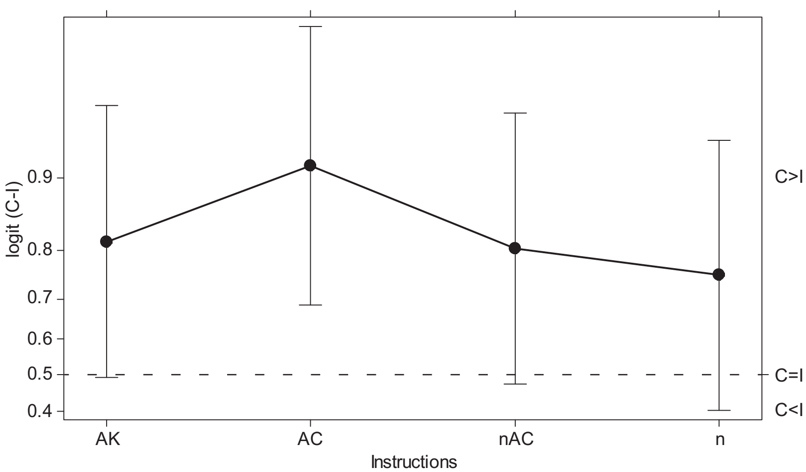


Figure 2. Plot of the effect of Instructions (emerging from the GLMM described in the main text of the paper) showing correct (C) vs. incorrect (I) solutions. Bars represent the 95% confidence interval.

numbers of correct solutions > numbers of incorrect solutions and logit values smaller than 0.5 indicate that numbers of correct solutions < numbers of incorrect solutions. Thus, one sees in Figure 2 that in all instruction conditions correct solutions were more frequent than incorrect solutions (the means of the effects are always >0.5). But, as the model confirmed, when participants were explicitly advised that thinking in terms of contraries might help them to find the solution (AC), the frequency of correct solutions was significantly higher as compared to when they were not given any advice (AC vs. n: $z = 2.596$, $p < 0.001$) or when they were not advised using contraries (AC vs. nAC: $z = 1.962$; $p < 0.05$). Groups advised to use contraries also tended to be correct more frequently than groups who had been advised to use previous knowledge (AC vs. AK: $z = 1.791$, $p = 0.07$). No differences were found between the other three conditions (AK vs. n: $z = 0.844$, $p = 0.398$; AK vs. nAC, $z = 0.178$, $p = 0.858$; n vs. nAC: $z = -0.667$, $p = 0.504$). Thus, giving explicit advice to use contraries to find a solution (AC) turned out to be the only instruction condition associated with an increase in the number of correct responses.

Perceptually based solutions vs. prior knowledge-based solutions. A second question regarded whether the frequency of solutions based on perceptual operations or prior geometrical knowledge differed in the four instruction conditions. In order to test this we recoded responses into a new binary dependent variable, perceptual solutions (PS) vs. prior knowledge solutions (KS)—independently of whether they were correct or incorrect solutions—and tested various GLMMs (link function: Logit).

As in the previous case, we first studied the effects of the two random effects (Groups and Problems) by testing whether the more complex model (with both random intercepts and random slopes, in addition to the Instruction fixed effect) differed in terms of fit from the simpler model which includes only the random intercepts (likewise in addition to the Instruction fixed effect). No significant difference emerged (chi-square = 3.416, df = 18, $p = 0.99$). AIC was smaller in the simpler model (AIC Instruction + random intercepts$_{\text{(Groups, Problems)}}$ = 506.33; AIC Instruction + random intercepts and slopes$_{\text{(Groups, Problems)}}$ = 538.91). Thus, only random intercepts were included for these two random effects in the next model tested, where the Instruction fixed effect was eliminated. The model including Instruction guaranteed a significantly better fit to the data as compared to the model including only random effects (chi = 18.349, df = 3, $p < 0.001$); Akaike's criterion (worse in the model without Instructions, AIC random intercepts = 518.6) also confirmed that the former minimises the loss of information from the data. In particular (see Figure 3), perceptual solutions were more frequent when contraries were used either explicitly (AC) or implicitly (nAC)—with no differences between these two conditions (AC vs. nAC;

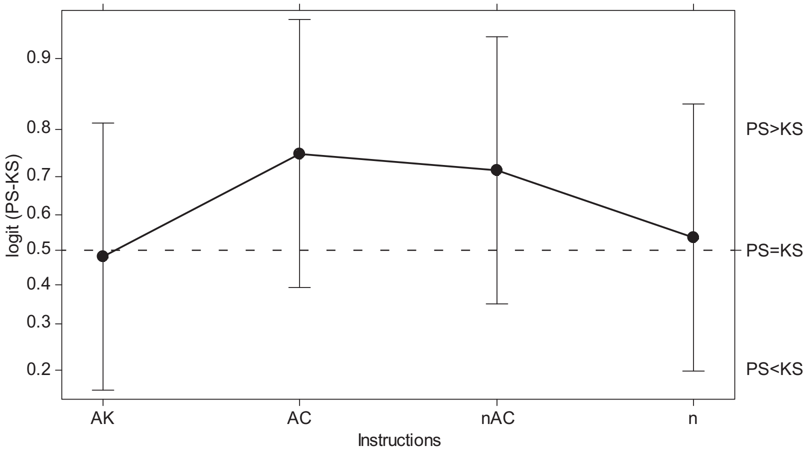


Figure 3. Plot of the effect of Instructions (emerging from the GLMM described in the main text of the paper) showing perceptual solutions vs. knowledge-based solutions (KS). Bars represent the 95% confidence interval.

$z = -0.604$, $p > 0.5$)—as compared to when participants were invited to use prior knowledge of formulas and rules (AK vs. AC: $z = -3.648$, $p < 0.0001$; AK vs. nAC: $z = -3.093$, $p = 0.002$) or when no advice regarding strategies was given (n vs. AC: $z = -3.008$, $p = 0.002$; n vs. nAC: $z = -2.436$, $p = 0.014$). Not giving any advice regarding strategy (n) or prompting participants to use previous knowledge (AK) led to similar results (n vs. AK: $z = -0.697$, $p = 0.485$). Given that in these two latter conditions we have logit values around 0.5 (AK = 0.481, $n = 0.536$), this also means that knowledge-based solutions were approximately as frequent as perceptual solutions in these two instruction conditions. Conversely, as the logit values in the other two instruction conditions indicate (AC = 0.751; nAC = 0.713), stimulating participants to explore the structure of the problem in terms of opposites, both when giving them hints and when not giving them hints about the utility of doing this, prompted them to seek perceptual solutions more often than solutions based on previously learnt knowledge.

Perceptual solutions: Incorrect vs. correct. So what happens if we specifically consider the distribution of incorrect perceptual solutions vs. correct perceptual solutions (cPS vs. iPS)?[5] Did differences emerge in the four instruction conditions? As previously, we first studied the random effects (Groups and Problems) by testing whether the more complex model which includes both

[5] It is statistically appropriate to parameterise a multinomial model as a series of binomial contrasts (Alison, 1984; Dobson & Barnett, 2008).

random intercepts and random slopes differed in terms of fit from the simpler model which includes only random intercepts, in both cases with the addition of the Instruction fixed effect. No significant differences emerged (chi-square = 1.234, df = 9, $p = 0.998$). Since AIC was smaller in the simpler model (AIC Instructions + random intercepts$_{(\text{Groups, Problems})}$ = 168.59; AIC Instructions + random intercepts and slopes$_{(\text{Groups, Problems})}$ = 185.35), only random intercepts were included in the model tested next where the Instruction fixed effect was deleted. No significant differences emerged between the model which also included Instructions as compared to the model with only random effects (chi = 3.299, df = 3, $p < 0.347$); Akaike's criterion was smaller in the latter model (AIC random intercepts$_{(\text{Groups, Problems})}$ = 165.89). Thus, the proportion of correct and incorrect perceptual solutions did not vary between the four instruction conditions.

Prior knowledge-based solutions: Incorrect vs. correct. Similar results were found when we considered the distribution of incorrect vs. correct knowledge-based solutions (iKS vs. cKS). No significant differences emerged between the two models used to study the random effects (Groups and Problems) in addition to the Instruction fixed effect (chi-square = 0.143, df = 9, $p = 1$). AIC was smaller in the simpler model (AIC Instructions + random intercepts$_{(\text{Groups, Problems})}$ = 145.35; AIC Instructions + random intercepts and slopes$_{(\text{Groups, Problems})}$ = 163.21). No significant differences emerged when Instruction was eliminated from the model including random intercepts (chi = 1.012, df = 3, $p = 0.798$); Akaike's criterion was also smaller in this model (AIC random intercepts$_{(\text{Groups, Problems})}$ = 140.36). Thus, the proportion of correct and incorrect knowledge-based solutions did not vary between the four instruction conditions.

Correct solutions: Perceptually based vs. based on prior geometrical knowledge. Finally, we tested whether the instructions affected the distribution of the two types of correct responses (correct solutions based on prior knowledge, cKS, vs. correct perceptual solutions, cPS). As before, we first studied the effects of the two random variables (Groups and Problems) by testing whether the more complex model (which includes both random intercepts and random slopes) differed in terms of fit from the simpler model which includes only random intercepts (in both cases, this was in addition to the Instruction fixed effect). No significant differences emerged (chi-square = 1.147, df = 9, $p = 0.998$) but AIC was smaller in the simpler model (AIC Instructions + random intercepts$_{(\text{Groups, Problems})}$ = 267.44; AIC Instructions + random intercepts and slopes$_{(\text{Groups, Problems})}$ = 284.29). Thus, only random intercepts were included for the two random effects in the model tested next, where the Instruction fixed effect was deleted. The elimination of the fixed effect led to a significant difference in the fit of the two models (chi = 16.914, df = 3, $p < 0.001$), indicating that Instruction affects the distribution of the two types of correct response. Akaike's criterion also confirmed that the model without Instruction was a worse fit (AIC random intercepts$_{(\text{Groups, Problems})}$ = 278.35). Having said that in all four conditions correct perceptual solutions were more frequent than correct solutions based on previous knowledge (see Figure 4: Logit values are always >0.5), the first type of response (cPS) was significantly more frequent when an explicit or implicit suggestion to use contraries was provided (AC, nAC) as compared to when participants were stimulated to rely on previously acquired geometrical knowledge (AK vs. AC: $z = -3.188$, $p = 0.001$; AK vs. nAC: $z = -3.519$, $p = 0.001$).

Solution latencies

By studying the times needed to reach a solution we aimed to understand whether the groups that made use of contraries (explicitly, AC, or implicitly, nAC) solved the problems faster than those that did not make use of this strategy or, conversely, whether solution latencies were shorter only when explicit advice was offered—both when related to contraries (AC) and when unrelated (AK).

The times needed to find a solution were measured by two independent judges. Measurements started at the point when participants finished reading the text of the problem and terminated when they moved on to the next problem. In the two conditions where participants were first asked to list all the contraries related to the problem (nAC and AC), the starting point was defined as when they announced that they were starting the second phase after having finished the preliminary phase, i.e., the point at which they started to look for a solution to the problem. Inter-rater agreement turned out to be excellent (intraclass correlation coefficient, ICC = 0.976; 0.971 < 95% confidence interval < 0.978). The following analyses were performed on a data matrix obtained by averaging the response times measured by the two raters.

We studied how solution times were distributed. A Box−Cox transformation suggested that data were distributed logarithmically (lambda = 0.038). We thus performed GLMMs on response times based on a Gaussian distribution and a log link function.

No differences emerged in the fit of the models including both random intercepts and random slopes for the two random effects (Groups and Problems) in addition to the fixed effect (Instruction) and the model including only the fixed effect and the two random intercepts (chi-square = 0.10, df = 18, $p = 1$). AIC was smaller in the simpler model (AIC Instructions + random intercepts$_{(\text{Groups, Problems})}$ = 5985.8; AIC Instructions + random intercepts and slopes$_{(\text{Groups, Problems})}$ = 6673.8). Thus, only random intercepts

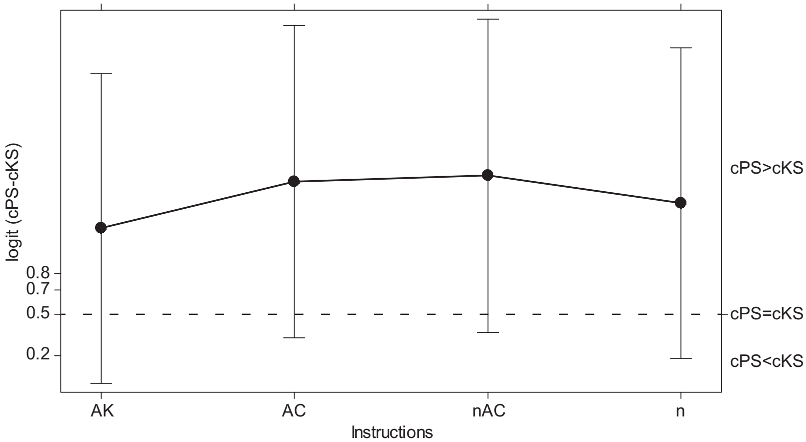


Figure 4. Plot of the effect of Instructions (emerging from the GLMM described in the main text of the paper) showing correct perceptual solutions (cPS) vs. correct knowledge-based solutions (cKS). Bars represent the 95% confidence interval.

were included in the model tested next, where the Instruction fixed effect was deleted. A significant difference was found (chi-square = 12.991, df = 3; $p =$ 0.004); this last model also had a worse fit to the data (AIC random $\text{intercepts}_{(\text{Groups, Problems})} = 5992.8$) In particular (see Figure 5), participants were faster when explicitly invited to use contraries as compared to when no

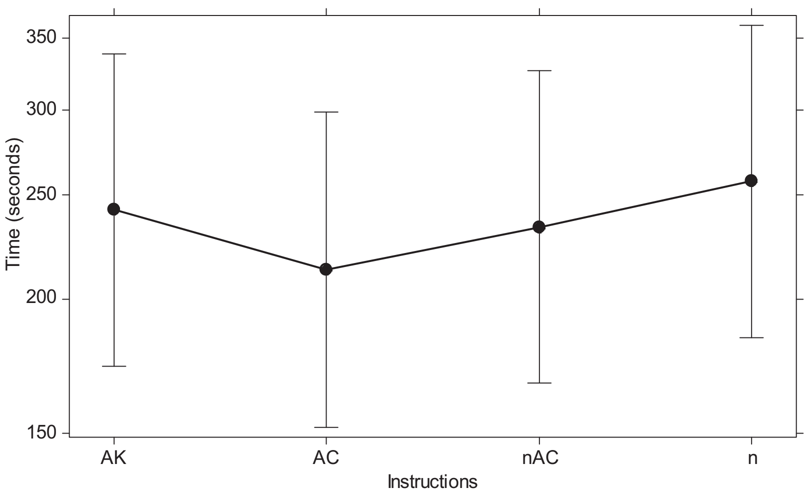


Figure 5. Plot of the effect of Instructions (emerging from the GLMM described in the main text of the paper) showing the time needed to find a solution. Bars represent the 95% confidence interval.

suggestion was provided (n vs. AC: $t = 3.525$, $p < 0.001$) or when they were invited to use previous geometrical knowledge (AK vs. AC: $t = 2.327$, $p < 0.05$). The implicit use of contraries (nAC) led to intermediate solution latencies (with no significant differences with any of the other three conditions).

Information focused on during the solution process

We finally tested whether differences emerged between the four instruction conditions in terms of the kind of information which participants focused on during the solution process. Dialogues between the participants were fully transcribed using a set of transcription conventions (Edwards, 1997; Gilbert, 1993; Potter, 1996) and coded based on an adaptation of the "thinking aloud coding scheme" (Hoffman & Lintern, 2006; Rasmussen et al., 1990). This adaptation was needed to make a good fit with the contents of the dialogues and the aspects of the process that we aimed to test. The coding table focused on two aspects:

Procedural aspects. We classified the chunks of dialogue where participants referred to what the problem asked them to do (*goal, G*); those where they referred to means, steps and information useful for reaching the solution (*means, M*); where they went back to the text of the problem and re-read it (*reading, RD*); where they paraphrased in their own words the text of the problem (*reformulation, RF*); where they applied true mathematical/geometrical formulas or self-instructions (these latter were either true or false but in any case they were dealt with as constraints to be followed) (*rules, R*); where they referred to aspects that might be derived from knowledge acquired in the past but that in any case referred to "visible" aspects (*visible rules, VR*) and where they referred to visual aspects, including descriptions of what was represented on the sheets of paper and any possible manipulations (*visual operations, VO*).

Features discussed. We also classified the chunks of dialogues where participants referred to the overall structure of the problem (*whole, W*); to a specific element of it (*part, P*); to the relation between a specific element and the overall structure of the problem (*whole–part, WP*); to the relation between two or more parts of the problem (*part–part, PP*) or to the relation between the overall structure of the problem and another hypothetical/imagined overall structure (*whole–whole, WW*).

Two independent judges did the codification (Cohen's K index $= 0.84$). We were interested in exploring whether the four instructions specifically increased or decreased the use of each of these aspects during the solution process. In order to avoid biases due to different overall numerosity, we transformed the raw frequencies into proportional data by dividing the counts for each category in a given group of participants analysing a given problem taking into account the total number of chunks produced by that group when analysing that problem (e.g., frequency of chunks referring to *visual rules* produced by group 1 when analysing the parallelogram problem in the AK condition/sum of the frequencies for all categories produced by group 1 when analysing the parallelogram problem in the AK condition). These weighted frequencies were then analysed using GLMMs, with Instruction as a fixed effect and Problems and Groups as random effects.[6]

The steps which were followed in the application of GLMMs to these data are the same as those used in the foregoing analyses: We first tried to ascertain which was the best way of modelling the random effects (Problems and Groups), i.e., whether only in terms of random intercepts or both random intercepts and slopes, and then tested whether the fit of the model worsened when the fixed effect (Instruction) was eliminated from the model. When no significant differences emerged, Akaike's criterion was used to identify the best model (smaller AIC).

Tables 2 and 3 summarise the outcomes of these analyses, for each individual category of information focused on. For the sake of simplicity, in column 2 only the difference between the two final models used to test the effect of Instructions is reported (i.e., the two best models in terms of the fit of the random components, with and without the fixed effect). When Instructions turned out to be critical (i.e., to guarantee a better fit), we analysed which out of the four instruction conditions differed (significant differences are reported in column 3 of Table 2; see also Figure 6).

The analyses revealed that Instructions affected procedural aspects (Table 2) while no systematic differences emerged concerning the type of features focused on during the solution process—i.e., whether individual aspects (P, W) or relational aspects (PP, WW, PW) (Table 3) were focused on. In particular, prompting participants to look for contraries (either explicitly, AC, or implicitly, nAC) led them to concentrate more frequently on what they were being asked to find (goal-directed behaviour). It also led them to expand on the problem by reformulating it (reformulation) as compared to simply re-reading it (reading was particularly frequent when no specific hints were provided) and to operate on visual aspects of the problem, e.g., by modifying orientation and localisation, separating and reorganising parts or the overall structure (visual operations). Finally, there were fewer references to mathematical/geometrical formulas (rules based on previous knowledge—but not visual rules—were instead significantly more frequent in AK).

[6] Crawley (2012) suggested modelling data related to proportions in the glmer R function following a binomial distribution since proportions vary between 0 and 1.

TABLE 2
Results of the GLMMs used to study the effects of the fixed factor Instruction on the proportional use (weighted frequencies) of various different procedural aspects related to the solution process

Aspect focused on [yes/no effect of Instruction]	*Compared models (best modelling of random effects, with and without the fixed factor)*	*Significant differences between the levels of Instructions*
Goal [yes]	Mod1 [Instruction + random intercepts$_{(Group,\ Problems)}$] vs. Mod2 [random intercepts$_{(Group,\ Problems)}$]: Chi = 7.556, df = 3, $p = 0.05$	AK vs. AC ($z = -1.885$, $p = 0.05$) n vs. AC ($z = -2.659$, $p = 0.0078$) n vs. nAC ($z = -1.839$, $p = 0.05$)
Means [no]	Mod1 [Instruction + random intercepts$_{(Group,\ Problems)}$ and random slope$_{(Problems)}$] vs. Mod2 [random intercepts$_{(Group,\ Problems)}$ and random slope$_{(Problems)}$]: Chi = 2.307, df = 3, $p = 0.5$ (AIC Mod2 < AIC Mod1)	
Reading [yes]	Mod1 [Instruction + random intercepts$_{(Group,\ Problems)}$] vs. Mod2 [random intercepts$_{(Group,\ Problems)}$]: Chi = 8.242, df = 3, $p < 0.05$	n vs. AC ($z = 2.570$, $p = 0.01$) n vs. nAC ($z = 2.480$, $p < 0.02$)
Reformulating [yes]	Mod1 [Instruction + random intercepts$_{(Group,\ Problems)}$] vs. Mod2 [random intercepts$_{(Group,\ Problems)}$]: Chi = 7.109, df = 3, $p < 0.05$	n vs. AK ($z = 2.709$, $p = 0.006$) AC vs. AK ($z = 1.744$, $p = 0.05$) nAC vs. AK ($z = 2.214$, $p = 0.026$)
Rules [yes]	Mod1 [Instruction + random intercepts$_{(Group,\ Problems)}$ and random slope$_{(Problems)}$] vs. Mod2 [random intercepts$_{(Group,\ Problems)}$ and random slope$_{(Problems)}$]: Chi = 7.926, df = 3, $p < 0.05$	nAC vs. AK ($z = -2.211$, $p < 0.05$) AC vs. AK ($z = -1.774$, $p = 0.05$)
Visual rules [no]	Mod1 [Instruction + random intercepts$_{(Group,\ Problems)}$] vs. Mod2 [random intercepts$_{(Group,\ Problems)}$]: Chi = 2.931, df = 3, $p = 0.4$ (AIC Mod2 < AIC Mod1)	
Visual operations [yes]	Mod1 [Instruction + random intercepts$_{(Group,\ Problems)}$] vs. Mod2 [random intercepts$_{(Group,\ Problems)}$]: Chi = 7.136, df = 3, $p < 0.05$	AC vs. n ($z = 1.712$, $p = 0.05$) nAC vs. n ($z = 2.418$, $p = 0.015$) nAC vs. AK ($z = 1.193$, $p < 0.05$)

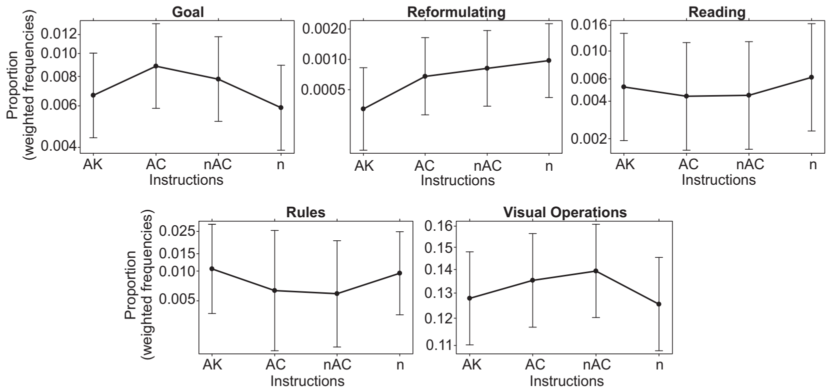


Figure 6. Plot of the effect of Instructions showing weighted frequencies of the procedural aspects which turned out to be significant in the GLMMs described in Table 2 (column 3). Bars represent the 95% confidence interval.

Some observations on the random factors

In this study various different geometrical problems were used. They shared the characteristics of being relatively doable problems which had already been studied (or at least discussed) by psychologists interested in problem solving. Differences between them were to be expected, although these were not “planned differences”, and in fact this “Problem” variable was a random

TABLE 3
Results of the GLMMs used to study the effects of the fixed factor Instruction on the proportional use (weighted frequencies) of various different types of features discussed during the solution process

P [no]	Mod1 [Instruction + random intercepts$_{(\text{Group, Problems})}$ and random slope$_{(\text{Problems})}$] vs. Mod2 [random intercepts$_{(\text{Group, Problems})}$ and random slope$_{(\text{Problems})}$]: Chi = 2.765, df = 3, $p = 0.429$ (AIC Mod2 < AIC Mod1)
W [no]	Mod1 [Instruction + random intercepts$_{(\text{Group, Problems})}$ and random slope$_{(\text{Problems})}$] vs. Mod2 [random intercepts$_{(\text{Group, Problems})}$ and random slope$_{(\text{Problems})}$]: Chi = 2.864, df = 3, $p = 0.413$ (AIC Mod2 < AIC Mod1)
PP [no]	Mod1 [Instruction + random intercepts$_{(\text{Group, Problems})}$ and random slope$_{(\text{Problems})}$] vs. Mod2 [random intercepts$_{(\text{Group, Problems})}$ and random slope$_{(\text{Problems})}$]: Chi = 3.483, df = 3, $p = 0.322$ (AIC Mod2 < AIC Mod1)
WP [no]	Mod1 [Instruction + random intercepts$_{(\text{Group, Problems})}$ and random slope$_{(\text{Problems})}$] vs. Mod2 [random intercepts$_{(\text{Group, Problems})}$ and random slope$_{(\text{Problems})}$]: Chi = 3.538, df = 3, $p = 0.315$ (AIC Mod2 < AIC Mod1)
WW [no]	Mod1 [Instruction + random intercepts$_{(\text{Group, Problems})}$] vs. Mod2 [random intercepts$_{(\text{Group, Problems})}$]: Chi = 1.979, df = 3, $p = 0.576$ (AIC Mod2 < AIC Mod1)

effect in our experimental design. Some differences did emerge: The frequency of correct solutions was particularly high for some problems (the nine dots and circumference problems) and particularly low for others (the ducks and square problems) while the number of perceptual solutions was particularly frequent for the nine dots problem and solutions based on previous knowledge were particularly frequent for the window problem. In addition, the time needed to find a solution was much longer for one problem (the square problem) as compared to the others. However, in the majority of cases (namely, in all the findings regarding the dependent "type of solution" and "solution latencies" dependent variables and for most of the findings concerning which procedural aspects were focused on during the solution process), when a significant effect of Instruction was found, intercepts were enough to model the random variability of the data related to the Problem factor (as well as the Group factor). The fit did not improve (on the contrary, it got worse) when random slopes were added. A random intercept model assumes that whatever the effect of Instructions is, it will be the same for all of the groups and all of the problems. Conversely, a random slope model assumes that the effect of Instructions might be different for each group and each problem, i.e., that Groups and Problems are not only allowed to have differing intercepts, but they are also allowed to have different slopes for the effect of Instructions. The fact that only random intercepts were needed thus indicates that the effect of Instructions was consistent across problems and groups. In other words, there was no evidence that the impact of Instructions on the dependent variables varied for each individual problem. Only when we considered the features focused on during the solution process (Table 3) was the fit of the random component better in the majority of cases with random slopes models than with only random intercepts models—this means that the instructions given affected some problems more than others. However, in these cases (Table 3) a general effect of Instructions was not found.

GENERAL DISCUSSION

In "Knowledge and Error" by Mach (1905), it is clearly stated that understanding what happens when we transform things into their opposites (e.g., compress–rarefy, cool down–warm up) and identifying the nature of the relationship between these phenomena may enable us to discover new physical laws. A more modest aim of this paper was to test whether stimulating the participants in our study to manipulate the characteristics of a problem in terms of contraries would affect how successful they were at solving geometrical problems. We were also interested in the impact of this manipulation on which aspects they focused on and on the time needed to reach a solution.

As the statistical analyses revealed, general effects of explicitly encouraging participants to reason in terms of contraries (AC) were found. A first effect was an increase in the number of groups who found the correct solution. A second effect was that participants were more likely to reason by referring to the perceptual structure of the problem (i.e., searching for a perceptual solution) rather than to previously learnt notions or rules. This was also reflected in the change in the type of information which they focused on, i.e., they referred less frequently to mathematical/geometrical formulas and more often relied on visual operations and reformulating the problem while bearing in mind the goal they needed to reach. In addition, this strategy also seemed to be associated with a reduction in the time needed to find a solution (solution latencies). We are not going to put particular emphasis on this last result for which further checks are needed, e.g., by adding an initial phase to the experimental design, similar to the one used here, but not involving contraries. At this point we prefer to concentrate on the other findings and make some considerations about those.

Easily applicable strategy

Elaborating on the structure of a problem in terms of contraries is an easily accessible strategy: It does not require expert competence since opposites are primal in human cognition in general and in spatial cognition in particular (as pointed out in the introduction to this paper). This simple strategy proved to enhance success.

Is this a generalisabile strategy? The next step would be to test whether this strategy would lead to similar positive results if we studied other types of problems i.e., not geometrical problems where the role of figural−perceptual processing is foremost. It has to be noted, however, that the problems considered in our study already represent an extension of the tasks typically used in previous studies investigating the effects of using contrast class cues, i.e., mainly Wason’s task (Gale & Ball, 2009). This stimulates us to ask whether there is a link between thinking in terms of opposites and “being creative” by which we mean being able to find alternative, divergent solutions rather than simply reproducing.

New insights into the relationship between thinking in contrasts and creativity

It has been suggested in the past that creativity is related to inductive reasoning (e.g., Guilford, 1964; Mumford, Connelly, Baugham, & Marks, 1994; Newell, Shaw, & Simon, 1962; Nickerson, 1999; Weisberg, 1999). In fact, induction and creativity are both characterised by the use of imagination and by the generation of new ideas (Johnson-Laird & Wason, 1977;

Nickerson, 1999) and they are both essential to hypothesis testing. Indeed hypothesis testing is an instance of inductive reasoning but it is at the same time an important component of creativity since the exploration of many scenarios and the generation of several hypotheses increase the likelihood that one will arrive at an original idea (Eysenck, 1995; Mednick, 1962).

It has been demonstrated that generating contrasting hypotheses necessitates the exploration of various alternatives (Gale & Ball, 2012; Oaksford & Charter, 1994)[7] and enhances creativity (Vartanian, Martindale, & Kwiatkowski, 2003). Attempts to directly connect the generation of contrasting hypotheses to divergent thinking are not new (Guilford, 1968, 1971) and techniques based on contraries to facilitate creative thinking have also been proposed. For example, in De Bono's lateral thinking approach (1967), he recommends not stopping at the first potential solution and he encourages people to continue to explore and search for alternative solutions by deliberately *reversing* the elements of a problem and the relationship between them. Kogan (1971) suggested *trying opposite means* to solve a problem when direct means fail. Zingales (1974) and Levine (1988) pointed out that a way to think creatively in problem solving consists of stretching a situation or its individual aspects up to extreme values using strategies which presuppose a variety of *oppositional processes* (exaggeration and diminution, enlargement and dissection or addition and subtraction) or choosing one of these strategies and then transforming the results obtained into the *opposite* (Zingales, 1974). New experimental evidence might refresh people's interest in understanding whether contraries play a role in problem solving and creative thinking and in discovering the cognitive basis of this link (taking advantage of new methodologies and new theoretical frameworks in psychology).

A strategy that remains anchored to the structure of the problem

To conclude, we would like to put forward a consideration regarding epistemology. Exploring a problem in terms of the contrasts embedded in its structure might not *necessarily* lead to a correct solution or make the search for it any quicker, but this investigation is invariably apposite. The properties which are manipulated might not be critical (and thus will not be of assistance in the search for a solution), but they are in any case related to the essence of the problem. In this sense any exploration of the contrasts which are integral to the structure of the problem complies with the requisites of "epistemic vigilance", defined as a filter mechanism acting on the

[7] But it is also a fundamental component of deductive reasoning in syllogisms (Evans, Handley, Harper, & Johnson-Laird, 1999; Johnson-Laird, 1983; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991).

information received in a communicative reasoning process to guarantee the reliability of the information processed (Mercier & Sperber, 2011; Sperber et al., 2010). In the problem solving strategy that we suggest, it is the phenomenal–perceptual structure of the problem itself that acts as a filter. It shows the gaps to be filled and the direction to follow. Moreover, humans tend to rely on what is evident (Brem & Rips, 2000) and problem solvers are inclined to trust what the structure of the problem itself suggests, even when this is in contrast with their beliefs. In other words, it is easy to justify and trust in a strategy which uses the phenomenal–perceptual structure of a problem (even in group settings) because it is based on observation.

Manuscript received 29 December 2013
Revised manuscript received 17 November 2014
Revised manuscript accepted 27 November 2014
First published online 30 December 2014

REFERENCES

Allison, P. (1984). *Event history analysis*. Newbury Park, CA: Sage.

Anolli, L., Antonietti, A., Cantoia, M., & Crisafulli, L. (2001). Accessing source information in analogical problem solving. *Quarterly Journal of Experimental Psychology, 54A*, 237–261. doi:10.1080/02724980042000093

Antonietti, A. (2001). *Analogical discoveries. Identifying similarities to solve problems*. Rome: Carocci.

Antonietti, A., Angelini, C., & Cerana, P. (1995). *L'intuizione visiva* [Visual intuition]. Milan: Franco Angeli.

Ball, L. J., Marsh, J. E., Litchfield, D., Cook, R. L, & Booth, N. (2014). When distraction helps: Evidence that concurrent articulation and irrelevant speech can facilitate insight problem solving. *Thinking & Reasoning, 20*, 76–96. doi: 10.1080/13546783.2014.934399

Ball, L. J., & Stevens, A. (2009). Evidence for a verbally-based analytic component to insight problem solving. In N. Taatgen & H. van Rijn (Eds.). *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1060–1065). Austin, TX: Cognitive Science Society.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014a). lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.1-7). Retrieved from http://CRAN.R-project.org/package=lme4

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014b). lme4: Linear mixed-effects models using Eigen and S4. ArXiv e-print; submitted to *Journal of Statistical Software*. Retrieved from http://arxiv.org/abs/1406.5823

Bianchi, I., Burro, R., Torquati, S., & Savardi, U. (2013). The middle of the road: Perceiving intermediates. *Acta Psychologica, 144*, 121–135. doi:10.1016/j.actpsy.2013.05.005

Bianchi, I., & Savardi, U. (2008). *The perception of contraries*. Milan: Franco Angeli.

Bianchi, I., Savardi, U., & Burro R. (2011). Perceptual ratings of opposite spatial properties: Do they lie on the same dimension. *Acta Psychologica, 138*, 405–418. doi:10.1016/j.actpsy.2011.08.003

Bianchi, I., Savardi, U., Burro, R., & Martelli, M.F. (2014). Doing the opposite to what another person is doing. *Acta Psychologica, 151*, 117–133. doi:10.1016/j.actpsy.2014.06.003

Bianchi, I., Savardi, U., & Kubovy, M. (2011). Dimensions and their poles: A metric and topological approach to opposites. *Language and Cognitive Processes, 26*, 1232–1265. doi:10.1080/01690965.2010.520943

Bozzi, P. (1978). L'interosservazione come metodo per la fenomenologia sperimentale [Interobservation as method for the experimental phenomenology]. *Giornale Italiano di Psicologia, 5*, 229–239.

Bozzi, P., & Martinuzzi, L. (1989). Un esperimento di interosservazione [An inter-observational experiment]. *Rivista di psicologia, 1*, 1–46.

Branchini, E., Burro, R., & Savardi, U. (2009). Contraries in productive thinking. In U. Savardi (Ed.), *The perception and cognition of contraries* (pp. 203–224). Milan: McGraw-Hill.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*, 573–604 doi:10.1207/s15516709cog2404_2

Casasola, M. (2008). The development of infants' spatial categories. *Current Directions in Psychological Science, 17*, 21–25. doi:10.1111/j.1467-8721.2008.00541.x

Casasola, M., Cohen, L. B., & Chiarello, E. (2003). Six-month-old infants' categorization of containment spatial relations. *Child Development, 74*, 679–693. doi:10.1111/1467-8624.00562

Crawley, M. (2012). *The R book*. Chichester: Wiley.

Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Crystall, D. (1985). *A dictionary of linguistics and phonetics*. Malden, MA: Blackwell.

De Bono, E. (1967). *The use of lateral thinking*. London: Cape.

Dobson, A. J., & Barnett, A. (2008). *An introduction to generalized linear models* (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC Texts in Statistical Science.

Duncker, K. (1926). A qualitative (experimental and theoretical) study in productive thinking (solving of comprehensible problems). *Pedagogical Seminary, 33*, 642–708. doi:10.1080/08856559.1926.10533052

Duncker, K. (1935). *Zur psychologie des productive denkes* [The psychology of productive thinking]. Berlin: Springer.

Duncker, K. (1945). On problem-solving. *Psychological Monographs, 58*(5), 1–113. doi:10.1037/h0093599

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics, 61*, 103–131. doi:10.1007/s10649-006-0400

Edwards, D. (1997). *Discourse and cognition*. London: Sage.

St Evans, J. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hove: Erlbaum.

St Evans, J. B. T. (2014). Reasoning, biases and dual processes: The lasting impact of Wason's (1960) paper on the 2-4-6 task. *Quarterly Journal of Experimental Psychology* (Hove). Advanced online publication. doi: 10.1080/17470218.2014.914547.

St. Evans, J. B. T., Handley, S. J., Harper, C. N. J., & Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1495–1513. doi:10.1037/0278-7393.25.6.1495

Eysenck, H. J. (1995). *Genius: The natural history of creativity*. Cambridge: Cambridge University Press.

Farris, H., & Revlin, R. (1989). Sensible reasoning in two tasks: Rule discovery and hypothesis evaluation. *Memory & Cognition, 17*, 221–232. doi:10.3758/BF03197071

Fleck, J. I., & Weisberg, R. W. (2013). Insight versus analysis: Evidence for diverse methods in problem solving. *Journal of Cognitive Psychology, 25*, 436–463. doi:10.1080/20445911.2013.779248

Gale, M., & Ball, L. J. (2003). Facilitated rule discovery in Wason's 2-4-6 task: The role of negative triples. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the Twenty-Fifth Annual*

Conference of the Cognitive Science Society (pp. 438–443). Boston, MA: Cognitive Science Society.

Gale, M., & Ball, L. J. (2006). Dual-goal facilitation in Wason's 2-4-6 task: What mediates successful rule discovery? *Quarterly Journal of Experimental Psychology, 59*, 873–885. doi:10.1080/02724980543000051

Gale, M., & Ball, L. J. (2009). Exploring the determinants of dual goal facilitation in a rule discovery task. *Thinking & Reasoning, 15*, 294–315. doi:10.1080/13546780903040666

Gale, M., & Ball, L. J. (2012). Contrast class cues and performance facilitation in a hypothesis-testing task: Evidence for an iterative counterfactual model. *Memory & Cognition, 40*, 408–419. doi:10.3758/s13421-011-0159-z

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.

Giaquinto, M. (2007). *Visual thinking in mathematics*. Oxford: Oxford University Press.

Gilbert, N. (1993). *Researching social life*. London: Sage.

Gilhooly, K. J., Fioratou, E., & Henretty, N. (2010). Verbalization and problem solving: Insight and spatial factors. *British Journal of Psychology, 101*, 81–93. doi:10.1348/000712609×422656

Gorgorio, N. (1998). Exploring the functionality of visual and non-visual strategies in solving rotation problems. *Educational Studies in Mathematics, 35*, 207–231. doi:10.1023/A:1003132603649

Gorman, M. E., Stafford, A., & Gorman, M. E. (1987). Disconfirmation and dual hypotheses on a more difficult version of Wason's 2–4–6 task. *Quarterly Journal of Experimental Psychology, 39A*, 1–28. doi:10.1080/02724988743000006

Guilford, J. P. (1964). Creative thinking and problem solving. *Education Digest, 29*, 21–31.

Guilford, J. P. (1968). *Intelligence, creativity and their educational implications*. San Diego, CA: R. R. Knapp.

Guilford, J. P. (1971). *The nature of human intelligence*. London: McGraw-Hill.

Gutiérrez, A. (1996). Children's ability for using different plane representations of space figures. In A. R. Batturo (Ed.), *New directions in geometry education* (pp. 33–42). Brisbane: Centre for Mathematics and Science Education, Q.U.T.

Gutiérrez, A., Pegg, J., & Lawrie, C. (2004). Characterization of students' reasoning and proof abilities in 3-dimensional geometry. *Proceedings of the 28th Conference of International Group for the Psychology of Mathematics Education*, *2*, 511–518. Retrieved from http://www.uv.es/gutierre/archivos1/textospdf/GutPeggLaw04.pdf

Harrower, M. R. (1932). Organization in higher mental processes. *Psychologische Forschung, 17*, 56–120.

Herbst, P. G. (2006). Teaching geometry with problems: Negotiating instructional situations and mathematical tasks. *Journal for Research in Mathematics Education, 37*, 313–347.

Hoffman, R. R., & Lintern, G. (2006). Eliciting and representing the knowledge of experts. In K. A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 203–222). New York, NY: Cambridge University Press.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University press.

Johnson-Laird, P. N. (1993). *Human and machine thinking*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition, 16*(1), 1–61. doi:10.1016/0010-0277(84)90035-0

Johnson-Laird, P. N., & Byrne, R. (1991). *Deduction*. Hove: Erlbaum.

Johnson-Laird, P. N., & Wason, P. C. (1977). Introduction to hypotheses and theories. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 257–273). Cambridge: Cambridge University Press.

Jones, K., & Bills, C. (1998). Visualisation, imagery, and the development of geometrical reasoning. *Proceedings of the British Society for Research into Learning Mathematics, 18*, 123–128.

Jones, S. (2002). *Antonymy: A corpus-based perspective*. London: Routledge.

Jones, S., Murphy, M. L., Paradis, C., & Willners, C. (2012). *Antonyms in English: Construals, constructions, and canonicity. Studies in English language*. Cambridge: Cambridge University Press.

Kanizsa, G. (1973). Il problem solving nella psicologia della gestalt [Problem solving in gestalt psychology]. In G. Mosconi & V. D'Urso (Eds.), *La soluzione dei problemi* [Problem solving] (pp. 35–86). Florence: Giunti Barbera.

Kareev, Y., Halberstadt, N., & Shafir, D. (1993). Improving performance and increasing the use of non-positive testing in a rule-discovery task. *Quarterly Journal of Experimental Psychology, 46A*, 729–742. doi:10.1080/14640749308401036

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228. doi:10.1037/0033-295X.94.2.211

Klayman, J., & Ha, Y.-W. (1989). Hypothesis testing in rule discovery: Strategy, structure and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 596–604. doi:10.1037/0278-7393.15.4.596

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition, 29*, 1000–1009. doi:10.3758/BF03195762

Kogan, N. (1971). Educational implications in cognitive styles. In G. S. Lesser (Ed.), *Psychology and educational practice* (pp. 242–292). Glenview, IL: Scott & Foresman.

Köhler, W. (1925). *The mentality of apes*. London: Kegan, Trench.

Köhler, W. (1969). *The task of gestalt psychology*. Princeton, NJ: Princeton University Press.

Kubovy, M. (2002). Phenomenology, cognitive. In L. Nadel (Eds.), *Encyclopedia of cognitive science* (pp. 579–586). London, UK: Nature Publishing Group.

Lakoff, G., & Núnez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books.

Levine, M. (1988). *Effective problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs, 54*, 1–95. doi:10.1037/h0093502

Luchins, A. S. (1946). Classroom Experiments on Mental Set. *The American Journal of Psychology, 59*, 295–298. doi:10.2307/1416894

Luchins, A. S., & Luchins, E. H. (1950). New experimental attempts at preventing mechanization in problem solving. *Journal of General Psychology, 42*, 279–297. doi: 10.1080/00221309.1950.9920160

Luchins, A. S., & Luchins, E. H. (1970a). *Wertheimer's seminars revisited problem solving and thinking* (Vol. I). New York, NY: State University of New York at Albany.

Luchins, A. S., & Luchins, E. H. (1970b). *Wertheimer's seminars revisited problem solving and thinking* (Vol. II). New York, NY: State University of New York at Albany.

Luchins, A. S., & Luchins, E. H. (1970c). *Wertheimer's seminars revisited problem solving and thinking* (Vol. III). New York, NY: State University of New York at Albany.

Luo, J., Niki, K., & Knoblich, G. (2006). Perceptual contributions to problem solving: Chunk decomposition of Chinese characters. *Brain Research Bulletin, 70*, 430–443. doi:10.1016/j.brainresbull.2006.07.005

Mach, E. (1905). *Erkenntnis und irrutum* [Knowledge and error]. Leipzing: Barth.

Maier, N. F. R. (1930). Reasoning in humans. On direction. *Journal of Comparative Psychology, 10*, 115–143. doi:10.1037/h0073232

Maier, N. F. R. (1931a). Reasoning and learning. *Psychological Review, 38*, 332–346. doi:10.1037/h0069991

Maier, N. F. R. (1931b). Reasoning in humans. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology, 12*, 181–194. doi:10.1037/h0071361
Maier, N. F. R. (1945). Reasoning in humans. The mechanisms of equivalent stimuli and of reasoning. *Journal of Experimental Psychology: General, 35*, 349–360. doi:10.1037/h0053960
Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69*, 220–232. doi:10.1037/h0048850
Mercier, H., & Sperber, D. (2011). Why do humans reasons? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57–111. doi:10.1017/S0140525×10000968
Mumford, M. D., Connelly, M. S., Baugham, W. A., & Marks, M. A. (1994). Creativity and problem solving: Cognition, adaptability, and wisdom. *Roeper Review, 16*, 241–246. doi:10.1080/02783199409553589
Murphy, L. (2003). *Semantic relations and lexicons: Antonyms, synonyms and other semantic paradigms*. Cambridge: Cambridge University Press.
Nerney, G. (1979). The gestalt of problem solving: An interpretation of Max Wertheimer's productive thinking. *Journal of Phenomenological Psychology, 10*, 56–80. doi:10.1163/156916279×00059
Newell, A., Shaw, J. C., & Simon, H. A. (1962). The process of creative thinking. In H. Gruber, G. Terrell, & M. Wertheimer (Eds.), *Contemporary approaches to creative thinking* (pp. 43–62). New York, NY: Atherton.
Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
Nickerson, R. S. (1999). Enhancing creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 392–430). New York, NY: Cambridge University Press.
Oaksford, M. (2002). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. *Thinking & Reasoning, 8*, 135–151. doi:10.1080/13546780143000170
Oaksford, M., & Chater, N. (1994). Another look at eliminative and enumerative behaviour in a conceptual task. *European Journal of Cognitive Psychology, 6*, 149–169. doi:10.1080/09541449408520141
Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 835–854. doi:10.1037/0278-7393.18.4.835
Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. Keane & K. Gilhooley (Eds.), *Advances in the psychology of thinking* (pp. 1–44). London: Harvester-Wheatsheaf.
Öllinger, M., Jones, G., & Knoblich, G. (2008). Investigating the effect of mental set on insight problem solving. *Experimental Psychology, 55*, 269–282. doi:10.1027/1618-3169.55.4.269
C. Paradis, J. Hudson, & U. Magnusson (Eds.) (2013). *The construal of spatial meaning: Windows into conceptual space*. Oxford: Oxford University Press.
Paradis, C., & Willners, C. (2011). Antonymy: From convention to meaning-making. *Review of Cognitive Linguistics, 9*, 367–391. doi: 10.1075/rcl.9.2.02par
Potter, J. (1996). *Representing reality*. London: Sage.
Rasmussen, J. (1985). The role of hierarchical knowledge representation in decision-making and system-management. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-15*, 234–243.
Rasmussen, J., Petersen, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering*. New York, NY: John Wiley.
Rasmussen, J., Petersen, A. M., & Schmidt, K. (1990). *Taxonomy for cognitive work analysis*. Denmark: Risø National Laboratory.
Savardi, U., & Bianchi, I. (2009). The spatial path to contrariety. In U. Savardi (Ed.), *The perception and cognition of contraries* (pp. 66–92). Milan: McGraw-Hill.

Savardi, U., Bianchi, I., & Bertamini, M. (2010). Naive prediction of orientation and motion in mirrors. From what we see to what we expect reflections to do. *Acta Psychologica, 134*, 1–15. doi:10.1016/j.actpsy.2009.11.008

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*, 359–393. doi:10.1111/j.1468-0017.2010.01394.x

Tukey, D. D. (1986). A philosophical and empirical analysis of subjects' modes of inquiry in Wason's 2–4–6 task. *Quarterly Journal of Experimental Psychology, 38A*, 5–33. doi:10.1080/14640748608401583

Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A., & Arkkelin, D. L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology, 32*, 109–123. doi:10.1080/00335558008248237

Unal, H., Jakubowski, E., & Corey, D. (2009). Differences in learning geometry among high and low spatial ability pre-service mathematics teachers. *International Journal of Mathematical Education in Science and Technology, 40*, 997–1012. doi:10.1080/00207390902912852

Vallée-Tourangeau, F., Austin, N. G., & Rankin, S. (1995). Inducing a rule in Wason's 2–4–6 task: A test of the information-quantity and goal-complementarity hypotheses. *Quarterly Journal of Experimental Psychology, 48A*, 895–914. doi:10.1080/14640749508401422

Vallée-Tourangeau, F., & Payton, T. (2008). Graphical representation fosters discovery in the 2-4-6 task. *Quarterly Journal of Experimental Psychology, 61*, 625–640. doi:10.1080/17470210701268039

Vartanian, O., Martindale, C., & Kwiatkowski, J. (2003). Creativity and inductive reasoning: The relationship between divergent thinking and performance on Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 56*, 641–655. doi:10.1080/02724980244000567

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129–140. doi:10.1080/17470216008416717

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.

Weisberg, R. W. (1999). Creativity and knowledge: A challenge to theories. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 226–250). New York, NY: Cambridge University Press.

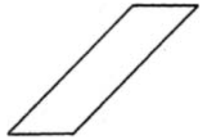
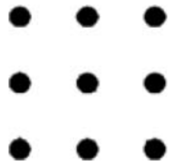
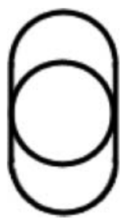
Wertheimer, M. (1945). *Productive thinking*. New York, NY: Harper. First appearance Über Schluzbprozesse in productiven Denken, 1919.

Wharton, C. M., Cheng, P. W., & Wickens, T. D. (1993). Hypothesis testing strategies: Why two goals are better than one. *Quarterly Journal of Experimental Psychology, 46A*, 743–758. doi:10.1080/14640749308401037

Wu, L., Knoblich, G., & Luo, J. (2013). The role of chunk tightness and chunk familiarity in problem solving: Evidence from ERPs and FMRI. *Human Brain Mapping, 34*, 1173–1186. doi:10.1002/hbm.21501

Zingales, M. (1974). *L'organizzazione della creatività* [The organization of creativity]. Bologna: Cappelli.

APPENDIX

Author/year	*Problem*	*Formulation*	*Figure*
Wertheimer (1945)	Parallelogram	The task is to find the formula to calculate the area of the figure and to demonstrate why it is correct.	
Maier (1930)	Nine dots	Four lines must be drawn between nine dots in such a way that all the dots are connected. The pencil must not be lifted from the paper and no line should be retraced.	
Harrower (1932)	Ducks	Swimming under a bridge there are two ducks in front of two ducks, two ducks behind two ducks, and two ducks in the middle. How many ducks are there in all?	No figure
Wertheimer (1945)	Altar window	Workers are painting and decorating the inner walls of a church. There is a circular window a little above the altar. As a decoration, the painters have been asked to draw two vertical lines tangentially to and of the same height as the circular window; they are then to add half circles above and below, closing in the figure. This area between the lines and the window is to be covered with gold. For every square inch, a certain amount of gold is needed. How much gold will be needed to cover this space (given the diameter of the circle); or, what is the area between the circle and the lines?	
Kanizsa (1973)	Square	Build a square by putting together six smaller figures: four right-angled isosceles triangles and two right-angled trapezoids having equal heights, but bases of different lengths.	

(*continued*)

Author/year	*Problem*	*Formulation*	*Figure*
Köhler (1969)	Circumference	Given a circle with a radius r, the task is to draw a rectangle in this circle. If we trace a line (l) inside the rectangle, what is its length?