Insight with Hands and Things

Frédéric Vallée-Tourangeau
Kingston University

Sune Vork Steffensen
University of Southern Denmark

Gaëlle Vallée-Tourangeau
Kingston University

Miroslav Sirota
University of Essex

Author Note

Correspondence concerning this article should be addressed to Frédéric Vallée-Tourangeau, Department of Psychology, Kingston University, Kingston-upon-Thames. UNITED KINGDOM, KT1 2EE, f.vallee-tourangeau@kingston.ac.uk tel: +44 (0)20 8417 2816, fax: +44 (0)20 8417 2388.

Abstract

Two experiments examined whether different task ecologies influenced insight problem solving. The 17 animals problem was employed, a pure insight problem. Its initial formulation encourages the application of a direct arithmetic solution, but its solution requires the spatial arrangement of sets involving some degree of overlap. Participants were randomly allocated to either a tablet condition where they could use a stylus and an electronic tablet to sketch a solution or a model building condition where participants were given material with which to build enclosures and figurines. In both experiments, participants were much more likely to develop a working solution in the model building condition. The difference in performance elicited by different task ecologies was unrelated to individual differences in working memory, actively open-minded thinking, or need for cognition (Experiment 1), although individual differences in creativity were correlated with problem solving success in Experiment 2. The discussion focuses on the implications of these findings for the prevailing metatheoretical commitment to methodological individualism that places the individual as the ontological locus of cognition.

Keywords: Problem solving, insight, task ecology, enactivism, methodological individualism

Insight with Hands and Things

1. Introduction

The psychology of problem solving has, over the years, split its research efforts tackling so-called analytic or transformation problems and insight problems. The former are well-defined problems with simple operators that can be applied to transform the initial problem presentation through a series of intermediate states—each intermediate state is a move in a logically specifiable problem space—to reach a desired configuration; the Tower of Hanoi or river crossing problems are good examples. In turn, insight problems are less well defined with no immediately obvious or effective operators that can be applied to transform the initial presentation into a solution. This is because insight problems are formulated in a manner that encourages a misleading interpretation and obscures a path to solution. For example, how can 17 animals be placed in four pens in such a manner that there is an odd number of animals in each pen? (to adapt a problem reported in Metcalfe and Wiebe, 1987). The problem masquerades as an arithmetic one, but an arithmetic solution is impossible (with whole animals/numbers); rather, a solution is possible when pens are projected as sets that can overlap.

The focus of the theoretical efforts for transformation problems is usually the effectiveness with which participants traverse the problem space, and performance is measured in the number and type of moves participants produce to reach the goal state. These efforts lend themselves to computational modelling of the move selection heuristics allegedly employed by participants. In turn, theoretical efforts for insight problem solving have focused on the processes that lead to a new interpretation, or restructuring of the problem representation that helps participants overcome an impasse and identify plausible solutions. The nature of the processes that result in insight has been the subject of some debate. One camp, inspired by Köhler's (1921/1957) ethnographic observations of the apparent suddenness of insight, suggest that largely unconscious and automatic processes evince a restructured mental representation of the problem—for example, Ohlsson's (1992) representational change theory and its more recent incarnation, redistribution theory (Ohlsson, 2011). Another camp holds that, like for transformation problems, insight solutions are distilled through conscious analytic processes that may or may not involve the restructuration of a mental representation (e.g., Fleck & Weisberg, 2004, 2013; Weisberg, 2015). There are two important features of the current debate about the mental processes implicated in insight problem solving. The first relates to the role of working memory; the second reflects a metatheoretical commitment to methodological individualism. Let's take each in turn.

If insight problem solving proceeds on the basis of a conscious analysis of the constituent elements of the problem and their relation, then one would expect measures of effortful cognitive analytical processing such as working memory capacity to be correlated with problem solving performance. On the other hand, if processing was largely unconscious, then working memory capacity might not be so relevant in the process of achieving insight. Using an individual differences approach, Gilhooly and Fioratou (2009) invited participants to solve series of insight and non-insight problems—from which composite performance scores were derived—and profiled their participants in term of verbal and visuo-spatial working memory using complex sentence, operation and visual pattern span tasks to determine the degree with which working memory measures correlated with the composite solution rate score for both types of problems. Verbal and visuo-spatial working memory span performance significantly predicted variance for *both* insight and non-insight problems. Gilhooly and Fioratou (2009) interpreted these findings in terms of the storage demands of keeping a rich problem representation in working memory such as to enhance the probability that “key elements (…) will be represented and available for reinterpretation” (p. 373). Working memory measures are strongly correlated with traditional measures of intelligence (e.g., Engle, Tuholski, Laughlin, & Conway, 1999; Oberauer, Schulze, Wilhelm, & Süß, 2005), and in turn measures of intelligence correlate with performance on insight problem solving (Davidson, 1995; Frederick, 2005). In like vein, research on individual differences in reasoning (e.g., Stanovich & West, 1998, Sirota, Juanchich, & Hagmayer, 2014) reveals that participants who score high on measures of intelligence, tend to engage in more rational thinking in a wide range of reasoning tasks. That research also implicates thinking dispositions in reasoning performance. Thus measures of open-minded thinking or willingness to engage in effortful thinking correlate with more rational thinking performance (Stavonich & West, 1998; Cacioppo, Petty, Feinstein, Jarvis, 1996).

A focus on cognitive capacities and thinking dispositions reflects a deep and pervasive commitment to methodological individualism, defined by Malafouris (2013, p. 25) as “the foregrounding of the human individual as the appropriate analytic unit and ontological locus of human cognition”. This commitment naturally encourages researchers to develop task procedures devoid of real-world meaning, goals and utilities, and that limit or prevent interactivity with the physical features of a problem, with the aim to identify ‘pure’ cognitive processes under controlled laboratory conditions (Vallée-Tourangeau & Vallée-Tourangeau, 2016). This commitment deflects attention away from the context of reasoning, and reinforces the focus on the capacities that an individual brings to a reasoning task. Yet, thinking and reasoning do not take place in a vacuum, and there is much evidence that systematic manipulations of task instructions, external representations, and artefacts can substantially transform deductive reasoning (e.g., Manktelow & Over, 1991), hypothesis-testing behavior (e.g., Gale & Ball, 2006; Vallée-Tourangeau & Payton, 2008; Vallée-Tourangeau, 2012), transformation problem solving (Zhang & Norman, 1994; Guthrie, Vallée-Tourangeau, Vallée-Tourangeau, & Howard, 2015), mental arithmetic (Carlson, Avraamides, Cary, & Strasberg, 2007; Lave, 1988; Vallée-Tourangeau, 2013), Bayesian reasoning (Vallée-Tourangeau, Abadie, & Vallée-Tourangeau, 2015) and insight problem solving (Weller, Villejoubert, & Vallée-Tourangeau, 2011). These context and representational effects encourage a transactional perspective on cognition. From this perspective, the cognitive capacities of the reasoner and the features of the context cannot be fruitfully segmented and their causal role defined orthogonally in the explanation of performance. A reasoner is embedded in a certain task environment that together configures a certain cognitive ecology within which certain cognitive abilities are manifested.

This transactional perspective encourages the exploration of the role of interactivity in problem solving (Steffensen, Vallée-Tourangeau, & Vallée-Tourangeau, 2016). In an interactive problem-solving environment, a problem is presented with manipulable constitutive elements. Take for example the matchstick arithmetic problems developed by Knoblich, Ohlsson, Haider, and Rhenius, (1999). The problems employ roman numerals in the shape of matchsticks that configure false arithmetic expressions (e.g., XI = III + III) that can be turned true by moving one matchstick (e.g., VI = III + III). However, in the original procedure employed by Knoblich et al. (and in their subsequent eye tracking experiment, Knoblich, Ohlsson, & Raney, 2001), the problems are presented on a computer screen and participants cannot manipulate the problem elements (and in the eye tracking experiment, even the participants' movements are constrained by the requirement of biting into a bar to stabilize the head and ensure more accurate eye tracking data). Participants stare at the computer display and mentally simulate matchstick movement; the perceptual feedback is invariant. Performance on this task is substantially transformed using a procedure wherein participants can manipulate the matchsticks (Weller et al., 2011). Moving a matchstick changes the physical appearance of the problem, prompts and guides new actions, and insight solutions are enacted through this dynamic cycle. Actions need not be premeditated; rather, simpler perception-action loops may shape, at different stages of the problem solving trajectory, the evolving physical configuration of the problem (Vallée-Tourangeau et al., 2015). An interactive problem solving environment foregrounds the importance of actions and the changes in action affordances wrought by the changes in the physical appearance of the problem. These reflections on research methodology, and the findings reported in Weller et al., suggest that the task ecology and the type of interactivity that it permits are important determinants of problem solving performance, above and beyond internal resources such as working memory capacity.

1.1 The Present Experiments

The primary aim of the experiments reported here was to determine whether different types of interactivity within different task ecologies influenced insight problem-solving performance. Both experiments employed the 17A problem, a pure insight problem according to the classification offered in Weisberg (1995). The 17A problem presents itself as involving an arithmetic solution yet this is only possible through the spatial arrangement of sets involving some degree of overlap (see Fig. 1). Two different task ecologies were created. In one, participants were given artefacts to build a model of the solution. They could not sketch a solution using a pen; only the material with which to build enclosures and 17 animal figurines were provided. In the second task environment, participants were invited to sketch a solution using a stylus and an electronic tablet. In that condition, no artefacts could be manipulated to spark ideas as participants drew their solution of the problem on the tablet.

We predicted that the type of interactivity—afforded by the task ecologies—would determine successful performance with the 17A problem. We expected a substantially higher rate of solutions in the model building condition, and this for two principal reasons. First, without the means to write down numbers and doodle various arithmetic operations, the focus on an arithmetic solution should more quickly dissipate in the model building than in the tablet condition. Second, building a model of the solution forces participants to tinker with the shape and spatial arrangement of the enclosures. Thus actions may enact a different path to solution, one that does not involve the brute labour of dividing 17 into 4 odd numbers.

To explore the importance of internal resources in problem solving, we also measured participants' cognitive capacities and thinking dispositions to determine whether they predicted problem-solving performance similarly in the two ecologies. Thus, in Experiment 1 we measured working memory capacity with two complex span tasks, an operation and a symmetry span, to establish participants' ability to maintain verbal and visuo-spatial material in working memory when concurrently performing simple arithmetic operations or symmetry judgments. If a change of representation is the product of an effortful cognitive process, as suggested by Kaplan and Simon (1990), measures of working memory capacity should predict performance on the 17A problem. In addition to cognitive capacities, we measured participants' disposition to engage actively in open-minded thinking (Baron, 1985)—i.e., the willingness to revise beliefs and abandon hypotheses in the face of contrary evidence—and need for cognition, participants' attitude toward and motivation for engaging in effortful thinking. These measures may predict how quickly participants abandon the unworkable application of a direct arithmetic solution, and hence their ability to formulate a solution involving overlapping sets. In Experiment 2 we sought to establish independently participants' creativity with the Biographical Inventory of Creative Behaviour (Batey, 2007) and the alternative uses task (Guilford, Christensen, Merrifield, & Wilson, 1978): more creative individuals may more often discover the overlapping-set solution. Finally, since the 17A problem is presented and initially interpreted as an arithmetic problem, Experiment 2 also profiled participants in terms of numeracy and mathematics anxiety. That is, more numerate participants who are less anxious about mathematics, may more quickly and more confidently realise that the solution to the problem cannot involve the division of an odd number into four odd numbers, without a remainder.

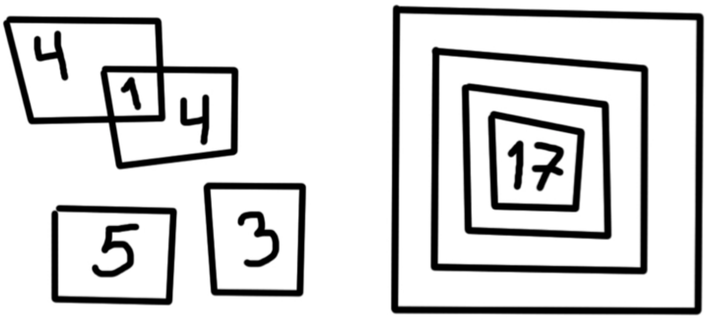


Figure 1. Possible solutions for the 17A problem.

2. EXPERIMENT 1

2.1 Method

2.1.1 Participants

Fifty psychology undergraduate and postgraduate students (44 females) received course credits for their participation (M_{age} = 24.2, SD_{age} = 8.1): Participants were randomly allocated to either the tablet condition (n = 24) or the model condition (n = 26). We used the sample sizes of prior studies where significant interactivity effects on problem-solving were demonstrated (e.g., Guthrie et al., 2015) to determine our stopping rule for the size of our sample. Such a sample size provides enough statistical power (i.e., 80%, assuming α = 5%, two-tailed test and unconditional exact test) to detect OR = 0.17 in success rate between the two conditions.

2.1.2 Procedure

2.1.2.1 Problem Solving Task. Participants were invited to solve the following problem: “How do you put animals in four enclosures in such a manner that there is an odd number of animals in each of the four pens?” All participants were first presented with a pen and blank sheet of paper and given three minutes to sketch possible solutions to the problem. No participant knew the solution to the problem or sketched overlapping pens during that initial period. After an interval of approximately 25 minutes—during which they completed a working memory test, see below—participants were allocated to either the tablet or the model condition and were given 10 additional minutes to solve the 17A problem. Participants in both conditions worked on the 17A problem on a table (118cm X 74cm) in an observation laboratory fitted with an overhead camera that recorded participants’ sketching and model building efforts.

2.1.2.1.1 Tablet. In this condition, participants were given a stylus and an electronic tablet (14.8cm X 19.7cm) with which to sketch a solution to the 17A problem; participants could draw and erase their workings with the stylus. The participants’ sketches were saved as MP4 video clips.

Figure 2. The ‘zebras’ used in the model condition of Experiment 1. The zebras were 3cm high, 2.5cm long, with a maximum width of 1cm.

2.1.2.1.2 Model. In this condition, participants were given approximately 20 pieces of pipe cleaners varying in length (short 20cm and long 30cm pieces) and 17 zebra paper clips (see Fig. 2). The pipe cleaning pieces were always placed to the left of the participants on the work table, the heap of 17 animals in front, and the problem statement to the right. Participants did not have a pen or piece of paper with which to sketch their solution; rather they had to build a model of the solution.

2.1.2.2 Individual Differences. Participants were profiled in terms of working memory capacity as well as their style of thinking and attitude towards thinking.

2.1.2.2.1 Working Memory. Participants completed two working memory tasks, an operation span and a symmetry span. Both span tasks, developed by the Georgia Tech Attention and Working memory lab (and used with permission), are automated tests, administered with E-Prime, each lasting 20-25 minutes—the timing of some of the elements in these tests is tailored to the average speed with which participants solve the arithmetic or symmetry problems during a practice session (Unsworth, Heitz, Schrock, & Engle, 2005).

In the operation span, participants are presented a series of letters to remember, each letter presentation is preceded by a simple arithmetic expression, such as (2*3) – 1 = ?, followed by a proposed answer which participants must judge as true or false. Letter series range in length between three and seven and three sets of each length is presented to participants. After each series, participants are shown a 4x3 array of letters—composed of letters that were part of the series and some that were not—and they click on each of the letter they remember as composing the series in the correct order of presentation. Participants' total number of correctly recalled letters over all series was used as the first measure of working memory capacity.

In turn, the symmetry span is an automated adaptation of the Corsi (1972) block test. Participants are presented with a series of coloured cells (or block) in a 4x4 matrix. Before the presentation of each block, participants are shown an 8x8 matrix with some cells coloured in black and are asked to determine whether the resulting pattern of black boxes is symmetrical along its vertical axis (see Unsworth et al., 2005). The series of coloured blocks varied in length between two and five, and participants were presented three sets of each length. After a set was presented, participants were prompted to click on a blank 4x4 grid the location of each coloured block and in the right order. The total number of blocks correctly recalled—out of a possible 42 across all sets—was the second measure of working memory capacity.

2.1.2.2.2 Need for Cognition. We adapted the short form of the need for cognition scale developed by Cacioppo, Petty and Kao (1984). We used the same 18 items but participants rated their endorsement with a 6-point scale (instead of a 5-point scale) anchored at 1 (completely false) and 6 (completely true) for items such as “I would prefer complex to simple problems”, “I really enjoy a task that involves coming up with new solutions to problems” and “Learning new ways to think doesn’t excite me very much” (this is a reverse coded item). The reliability of the Need for Cognition scale was excellent with a Cronbach’s alpha of .89.

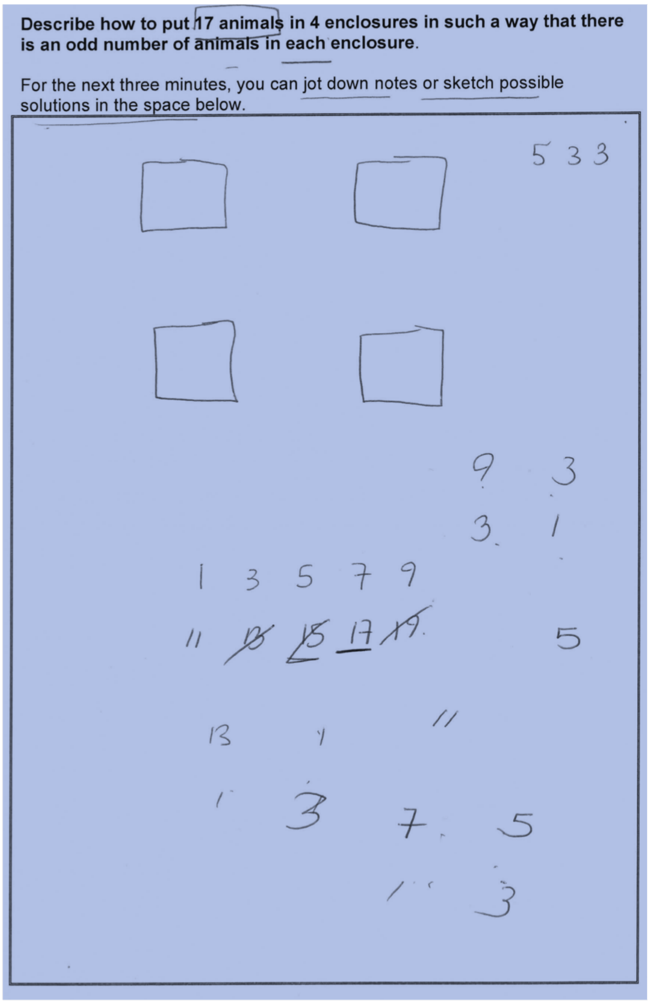
2.1.2.2.3 Actively Open-minded Thinking. We used the 7-item actively open-minded thinking scale reported in Haran, Ritov and Mellers (2013). Using a 7-point scale (1 = completely disagree, 7 = completely agree) participants rated their agreement with statements such as “People should revise their beliefs in response to new information or evidence” and “It is important to persevere in your beliefs even when evidence is brought to bear against them" (this is a reverse coded item). The reliability of the AOT scale was relatively poor with a Cronbach's alpha of .51.

The experimental session began with either the operation or symmetry span task (counterbalanced across participants), then the AOT and Need for Cognition scales (their order also counterbalanced), the three-minute preliminary problem solving period, the second working memory span task, and then the 10-minute problem solving period.

2.2 Results

2.2.1 Problem Solving Performance

None of the participants solved the problem during the initial three-minute period. All participants sketched answers (see Fig. 3 for an illustration) that clearly illustrated an interpretation of the problem as requiring an arithmetic solution. After a 25-min working memory assessment interval, participants were given an additional 10 minutes to solve the problem, either using a stylus and an electronic tablet to draw a solution, or using pipe cleaners and 'zebras' to build a model of the solution. Of the 24 participants in the tablet condition, none solved the problem in the 10-minute period (see Fig. 4, Panel B, for illustrations of the participants' sketching). That is, participants worked for the entire 10-minute period on discovering how an odd number could be split into 4 odd quantities, a mathematical impossibility with natural numbers. Thus participants in the tablet condition were never able to abandon the original problem interpretation and attempted to discover an arithmetic solution. Having said this, two participants in the tablet condition, at one point in their work, drew something that resembled overlapping sets but nonetheless subsequently laboured a direct arithmetic solution.



Describe how to put 17 animals in 4 enclosures in such a way that there is an odd number of animals in each enclosure.

For the next three minutes, you can jot down notes or sketch possible solutions in the space below.

Figure 3. Impasse measured during an initial 3-minute attempted solution to the 17A problem in Experiment 1. No participant in either the tablet or the model group sketched a solution to the problem during this initial period.

Of the 26 participants in the model condition, three systematically clipped the zebras onto the pipe cleaners during the 10-minute problem solving period. This was indeed an affordance of the artefacts employed in the model condition, but an unforeseen one when the material was initially piloted. In effect, by clipping the zebras onto the pipe cleaners, these participants could never discover the solution to the problem, since an animal could not be placed into more than one pen simultaneously. This type of problem solving trajectory would not have been possible had we chosen any other type of non-clipping figurines to correspond to the ‘animals’ in the problems. As a result,

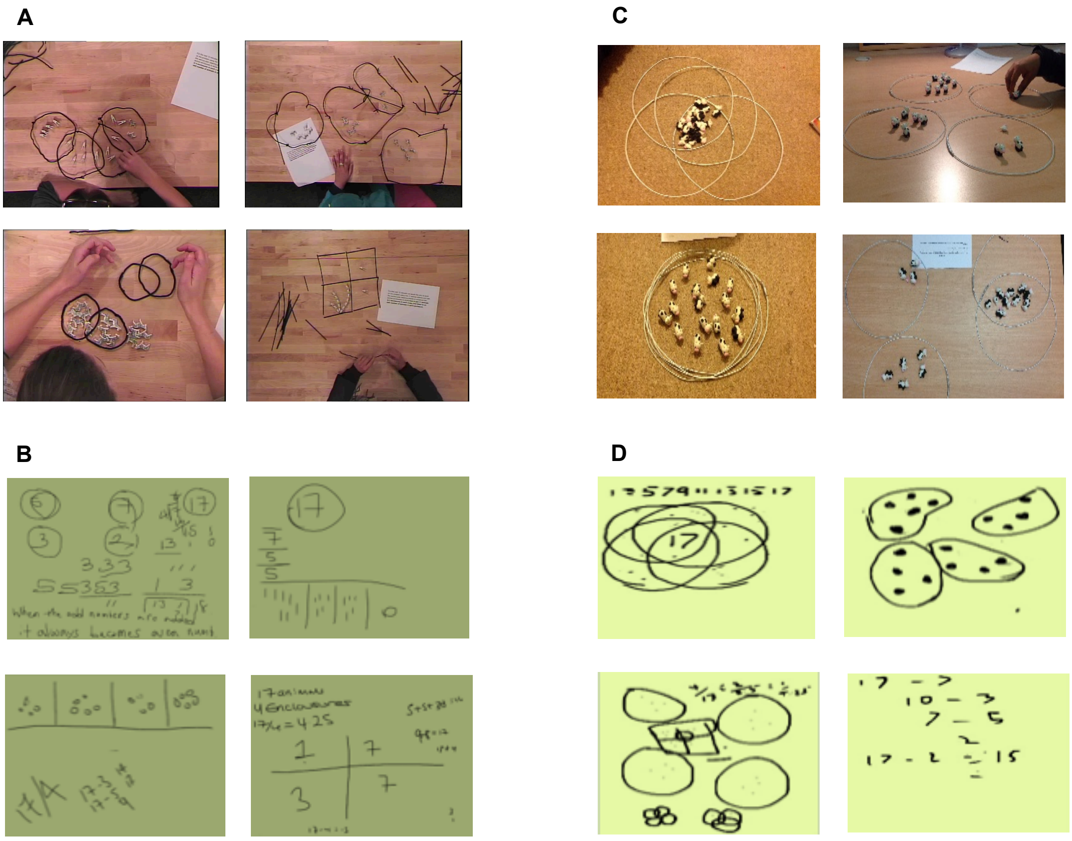


Figure 4. Examples of participants' solutions, partial solutions, and incomplete solutions after working for 10 minutes on the 17A problem in the model (A) and tablet (B) conditions of Experiment 1 and in the model (C) and tablet (D) conditions of Experiment 2.

we chose to remove these three participants from all subsequent analyses. Of the remaining 23 participants, 6 solved the problem outright (see Fig. 4, Panel A) and 4 offered partial solutions—that is solutions with overlapping sets, but ones for which a set intersection is taken as a separate pen, and while there is an odd number of animals in each resulting enclosure, this results in a five-pen solution. Of the 13 who did not solve the problem, 3 worked with overlapping sets but were unable to arrange the animals in a correct manner, and 10 build enclosures that never overlapped.

To avoid the issues with an empty cell we used a Barnard's exact unconditional test instead of the chi-square test; Barnard's test is considered to be a more powerful and less constrained alternative to the Fisher's exact test (e.g., Lydersen, Fagerlans & Laake, 2009). Thus 10 participants provided full or partial solutions to the problem in the model condition, compared to none in the tablet condition, a significant difference using the Barnard's exact unconditional test (3.64, *p* < .001).

Table 1

Mean Scores (and SD) on the Actively Open-minded Thinking (AOT) and Need for Cognition (nCog) scales, and Mean (and SD) Working Memory Capacity Scores, as Assessed with an Operation Span (O-Span) and a Symmetry Span (S-Span), for Solvers and Non-solvers in the Tablet and Model Groups of Experiment 1. t Values Refer to Group Differences (df = 45).

	Tablet		Model				
	M	*SD*	*M*	*SD*	*t*	*p*	Cohen's *d*
AOT							
Solvers	-	-	34.2	5.0			
Non-solvers	32.1	4.3	30.5	5.6			
Overall	**32.1**	**4.3**	**32.1**	**5.5**	**-0.004**	**.997**	**0.001**
nCog							
Solvers	-	-	75.3	14.2			
Non-solvers	73.4	13.0	75.7	14.0			
Overall	**73.4**	**13.0**	**75.5**	**13.8**	**-0.539**	**.593**	**0.160**
O-Span							
Solvers	-	-	51.5	7.5			
Non-solvers	53.5	11.1	53.5	10.4			
Overall	**53.5**	**11.1**	**52.7**	**9.1**	**0.286**	**.775**	**0.085**
S-Span							
Solvers	-	-	26.8	9.2			
Non-solvers	26.2	7.5	26.8	9.7			
Overall	**26.2**	**7.5**	**26.8**	**9.3**	**-0.268**	**.790**	**0.079**

Note. There were no solvers in the tablet group; the 'solvers' in the model group include the 6 correct solvers and the 4 partial solvers.

2.2.2 Individual Differences

Participants' scores on the Actively Open-minded Thinking (AOT) tests, the Need for Cognition scale, as well as on the two working memory tests—

the Operation and Symmetry Spans—are reported in Table 1. Participants in both groups scored the same on these measures—as we would expect with the random participant allocation—and there were no significant differences between groups; the largest difference was observed for the Need for Cognition scores, $t(45) = -.539$, $p = .593$. Thus, problem solving performance differences between groups cannot be attributed to differences along these thinking and working memory dimensions. Finally, we compared solvers and non-solvers in the model group. Here too success at solving the problem was unrelated to these individual differences, with the largest non-significant difference observed with the Actively Open-minded Thinking scale, $t(21) = -1.627$, $p = .119$.

2.3 Discussion

The 17A problem is a hard problem. After working on the problem for a total of 13 minutes—the initial 3-min sketching phase plus the 10 minutes in either one of the two experimental conditions—none of the participants in the tablet condition formulated a solution, while 43% built overlapping sets that led to a full or partial solution in the model building condition. On the basis of the participants' preliminary sketches, all interpreted the problem as one involving the direct application of arithmetic operators to yield a solution. Participants' arithmetic focus was unwavering in the tablet condition: they primarily worked on dividing 17 into four odd numbers, often by listing odd numbers from 1 to 17 and attempting to combine four that would tally up to 17. If pens were drawn, they were rarely modified once sketched, and remained as an unchanging and non-overlapping configuration of four separate areas throughout the problem solving session. In contrast, in the model condition, participants' attention and actions were directed, from the start, on the pens themselves. As a result, the spatial configuration of the pens, rather than quixotic attempts to divide 17 into four odd numbers, was for some participants the focus of their problem solving efforts. Thus, it is through the manufacturing and manipulation of the pens that some participants overcame the impasse: the action of building and re-building pens, especially in the absence of a stylus with which to sketch arithmetic operations, helped some participants to relax the initial focus on a direct arithmetic interpretation of the problem. Problem solving success was driven by the action repertoire afforded by the task environment. With pipe cleaners, participants worked on the pens, and some managed to create overlapping structures which enabled them to work out the distribution of individual animals. With a stylus, participants tackled the problem as in a more traditional pen and paper scholastic maths exercise, never achieving the required number-enclosure figure-ground inversion (see also Steffensen et al., 2016), and enclosures always remained an invariant perceptual background over which numbers were manipulated and superimposed.

Participants in the first experiment were profiled in terms of their dispositions to be actively open-minded thinkers and enjoyment in poring over difficult problems. Scores on neither dimension differentiated participants in the tablet and model building conditions, nor between those who formulated an overlapping solution and those who did not in the model building condition. In terms of working memory capacity, participants in the two groups did not differ either on their scores for the operation or the symmetry span. Similarly, no significant differences were observed in terms of working memory capacity between solvers and non-solvers in the model building condition. Problem solving success was determined by the task environment, the range of actions it afforded and the dynamic changes to the physical problem presentation evinced by these actions.

In the second experiment we aimed to replicate these findings using different artefacts and explore additional individual differences. One can hypothesize that a number of material properties determine the trajectory of insight: Pens can be formed with pieces of varying length, height, weight, rigidity and adhesion. They can even be formed by pre-given surfaces in different shapes and sizes. We conjectured that overlap-ability might be more easily enacted with hoops. In addition, overlapping hoops or circles are commonly encountered in the teaching of algebra of sets, as well as having common cultural symbolic currency (e.g., the Olympic rings, the front grille adornment of a German luxury car). The second experiment also avoided using artefacts for animals that could clip onto the perimeters of the pens. We improved the experimental procedure by showing participants allocated to the tablet condition a picture of the hoops and animal figurines before they sketched an answer to the problem. Thus the perceptual starting point was the same in both the tablet and model building conditions.

In the second experiment we explored other individual differences. First we sought to establish whether independent measures of creativity could predict problem solving success. Second, because the problem is initially interpreted as requiring the application of a direct arithmetic solution, participants who are less numerate or experience mathematics anxiety might not be sufficiently confident to discard the unworkable arithmetic strategy. We thus assessed the levels of subjective numeracy and maths anxiety. Finally, we also measured impulsivity: We conjectured that participants who scored high on impulsivity might more quickly abandon a certain problem solving trajectory and try something new, which may eventuate in loosening the focus on the arithmetic interpretation of the problem.

3. EXPERIMENT 2

3.1 Method

3.1.1 Participants

Forty-seven undergraduate and postgraduate students (38 females) received course credit for their participation (M_{age} = 21.5, *SD* = 4.73). Participants were randomly allocated to either the tablet (*n* = 23) or model (*n* = 24) condition. We used the same rationale as for Experiment 1 to determine our stopping rule for the sample size.

3.1.2 Procedure

Participants worked on the 17A problem using the same problem statement as in Experiment 1. However, unlike in Experiment 1, participants were not given pen and paper to sketch a solution during a three-minute preliminary phase. As in Experiment 1, participants in the tablet condition were given a stylus and an electronic tablet to sketch a solution. In the model building condition, participants were given 4 metal hoops (each 20-cm in diameter) and 17 animal figurines. Figure 5 illustrates the material employed in the model condition. Note that this image was also shown to the participants in the tablet condition.

3.1.2.1 Individual Differences. Participants were profiled in terms of their creativity, impulsivity, numeracy, and mathematics anxiety.

3.1.2.1.1 Biographical Inventory of Creative Behaviours (Batey, 2007). Participants completed a checklist of 34 different creative behaviours they might have engaged in during the past 12 months, such as “draw a cartoon”, “started a club, association or group”, “composed a poem”. Scores reflected the number of items ticked.

Figure 5. Material in the model condition of Experiment 2. The hoops had a diameter of 28cm. The animals were ‘cows’, 2.5cm high, 3.5cm long and 2cm wide.

3.1.2.1.2 Alternative Use Task (Guilford, Christensen, Merrifield, & Wilson, 1978). Participants were asked to generate as many alternative uses of a brick during a three-minute period. We only scored fluency based on how many unique uses were generated.

3.1.2.1.3 Barratt Impulsiveness Scale (BIS; Patton, Stanford, & Barratt, 1995). Participants completed the 30-item BIS, using a 6-point scale (1 = “Never”, 6 = “Always”) to answer items such as “I do things without thinking”, “I have racing thoughts”, “I concentrate easily” (reverse coded). The scale’s reliability was acceptable: Cronbach’s alpha = .78

3.1.2.1.4 Numeracy. Numeracy was measured using the subjective numeracy scale developed by Fagerlin, Zikmund-Fisher, Ubel, Jankovic, Derry, and Smith (2007). This scale is composed of eight items (such as “how good are you at calculating a 15% tip”). Participants answer using a 6-point scale (1 = “not good at all” and 6 “extremely good”). The sum ratings was the subjective numeracy score. Fagerlin et al. report strong positive correlations with this subjective measure and objective numeracy measured with series of arithmetic problems. The scale showed good reliability with Cronbach’s alpha at .84.

3.1.2.1.5 Mathematics Anxiety. Participants completed a 25-item Mathematics Anxiety Scale-UK (MAS-UK; Hunt, Clark-Carter & Sheffield, 2011). The questionnaire invited participants to imagine how anxious they would feel in certain situations (1 = “not at all” and 5 = “very much”), such as “Working out how much your shopping bill comes to” or “Taking a maths exam”. The scale had excellent reliability: Cronbach’s alpha = .92.

Three separate orders of the different scales were created to mitigate order effects with two constraints: (1) the 17A problem was always presented as the second task and (2) the numeracy and maths anxiety scales were always presented at the end of the experimental session.

3.2 Results

3.2.1 Problem Solving Performance

Four participants (or 17%) solved the 17A problem in the tablet condition, while 13 (or 54%) did in the model building condition, a significant difference, $\chi^2(1, N = 47) = 6.88$, $p = .009$, Cramer’s $V = .383$. Compared to Experiment 1, significantly more participants produced a correct solution to the problem in Experiment 2, $\chi^2(1, N = 94) = 6.97$, $p = .008$, Cramer's $V = .272$.

Table 2

Mean Scores (and SD) on the Biographical Inventory of Creative Behaviours (BCIB), the Alternative Uses Task (AUT), the Barratt Impulsiveness Scale (BIS), the Subjective Numeracy Scale (SNS), and the Mathematics Anxiety Scale (MAS) for Solvers and Non-solvers in the Tablet and Model Groups of Experiment 2. t Values Refer to Group Differences (df = 45).

	Tablet		Model				
	M	*SD*	*M*	*SD*	*t*	*p*	Cohen's *d*
BICB							
Solvers	11.0	5.7	9.7	4.8			
Non-solvers	6.9	4.0	6.1	2.3			
Overall	**7.6**	**4.5**	**8.0**	**4.1**	**-0.345**	**.732**	**0.103**
AUT							
Solvers	7.3	2.6	6.4	3.2			
Non-solvers	4.5	2.2	3.6	2.6			
Overall	**5.0**	**2.4**	**5.1**	**3.0**	**-0.156**	**.877**	**0.046**
BIS							
Solvers	96.5	20.9	100.3	14.3			
Non-solvers	94.7	9.6	101.8	16.9			
Overall	**94.8**	**11.6**	**101.0**	**15.8**	**-1.522**	**.135**	**0.454**
SNS							
Solvers	37.5	9.1	29.7	9.4			
Non-solvers	28.7	7.0	25.9	8.0			
Overall	**30.3**	**7.9**	**28.0**	**8.5**	**0.960**	**.342**	**0.286**
MAS							
Solvers	45.5	5.4	52.2	18.8			
Non-solvers	54.2	11.6	64.8	22.0			
Overall	**52.7**	**11.2**	**58.0**	**20.5**	**-1.086**	**.283**	**0.323**

Note. There were 4 solvers in the tablet group and 13 in the model group.

3.2.2 Individual Differences

The mean scores on the measures of individual differences are reported in Table 2. Overall, the groups did not significantly differ on any of the five dimensions; largest non-significant difference was with the BIS scores, $t(45) = -1.52$, $p = .135$. Table 3 reports the correlations among the different measures of individual differences as well as the point-biserial correlations with problem solving success in the tablet and model conditions. We note the negative correlations between measures of numeracy (SNS) and mathematics anxiety (MAS) in the tablet, $r(21) = -.75$, $p < .001$, and model condition, $r(22) = -.409$, $p = .047$. Impulsiveness (BIS) was positively correlated with maths anxiety in the model condition, $r(22) = .59$, $p = .002$, although not in the tablet condition, $r(21) = .08$, $p = .724$.

Of more interest are the point-biserial correlations with problem solving success. Measures of creativity were positively correlated with success, particularly output on the Alternative Uses Task in the tablet, $r_{pb}(21) = .43$, $p = .040$, and model condition, $r_{pb}(22) = .47$, $p = .022$. Scores on the Biographical Inventory of Creative Behaviours were positively correlated with success in the model condition, $r_{pb}(22) = .45$, $p = .029$, although not significantly so in the tablet condition, $r_{pb}(21) = .35$, $p = .097$. Finally, numeracy was positively correlated with success in the tablet, $r_{pb}(21) = .43$, $p = .041$, but not in the model condition, $r_{pb}(22) = .23$, $p = .287$.

3.3 Discussion

The second experiment sought to establish the importance of the artefacts employed in determining problem solving success as well as whether measures of creativity, impulsiveness, numeracy and mathematics anxiety correlated with discovering the solution to the 17A problem. As in Experiment 1, participants were more likely to build an overlapping set solution in the model condition than to sketch one in the tablet condition. In addition, a significantly greater number of participants discovered the solution in the second experiment than in the first experiment. Participants’ experience

Table 3.

Intercorrelations for Problem Solving Success and Measures of Creativity (BCIB, AUT), Impulsiveness (BIS), Numeracy (SNS) and Mathematics Anxiety (MAS) in the Tablet (Left Panel) and Model Groups (Right Panel).

	Tablet Group							**Model Group**					
	1	2	3	4	5	6		1	2	3	4	5	6
1 **BCIB**	--						1 **BCIB**	--					
2 **AUT**	.141	--					2 **AUT**	.172	--				
3 **BIS**	.124	-.180	--				3 **BIS**	-.115	-.319	--			
4 **SNS**	-.040	.244	-.316	--			4 **SNS**	.370	.070	.079	--		
5 **MAS**	-.127	-.202	.078	-.749 **	--		5 **MAS**	-.346	-.397	.593 **	-.409 *	--	
6 **Success**	.354	.431 *	.027	.429 *	-.302	--	6 **Success**	.445 *	.467 *	-.049	.227	-.314	--

Note. **BICB** = Biographical Inventory of Creative Behaviours; **AUT** = Alternative Uses Task; **BIS** = Barratt Impulsiveness Scale; **SNS** = Subjective Numeracy Scale; **MAS** = Mathematics Anxiety Scale. The correlations with Success are point-biserial correlations. * $p < .05$ ** $p < .01$

with hoops and their commonly encountered arrangements outside the laboratory might have contributed to the enhanced solution rate in Experiment 2. Furthermore, since participants did not have to build the enclosures, relatively more time could be devoted to tinkering with set arrangements than in Experiment 1.

A number of material parameters may determine the problem solving trajectory. The model condition can be designed in a number of different ways: Pens can be constructed with pieces of varying length, weights, and rigidity. The resulting pens can have a certain height and weight. Pens can be presented as ready-made forms—such as the hoops in Experiment 2—which can also vary in shape and size. Variations of the figurines and work space are also possible. Our aim was not to determine the optimal physical parameters that lead to the highest rate success in the 17A problem. Rather, unlike current models of problem solving, such as Ohlsson’s (2011) redistribution theory or Fleck and Weisberg’s (2013) integrated framework, an ecological perspective encourages that kind of analysis and helps understand differences in solution rates as a function of the physical parameters of the thinking environment.

While task ecology had a substantial impact on problem solving success in Experiment 2, independent measures of creativity were associated with solving the 17A problem. In both the tablet and model condition, higher response outputs in the Alternative Uses Task (AUT) was associated with producing a correct solution, and higher scores on the Biographical Inventory of Creative Behaviours (BICB) was correlated with success in the model but not in the tablet condition. It is interesting to note that while scores on the

Biographical Inventory of Creative Behaviours were positively correlated with response fluency in the Alternative Uses Task, these correlations were not significant. This is perhaps not surprising. The BICB records protracted activities that are deemed creative. These activities also require effort, commitment, motivation and resources. In turn, the AUT reflects an agility of mind that drives the generation and simulation of alternative uses of a common object in a very short time span. Such ability may not be necessarily implicated in bringing to completion acts such as “starting a club”, “making someone a present” or “acting in a dramatic production”, which are among the items that compose the BICB. In turn a high response fluency may reflect a participant’s drive to explore different spatial configurations in the 17A problem. While we conjectured that measures of impulsivity might correlate with success at the task, scores on the BIS and solving the 17A problem were unrelated. Finally, higher levels of subjective numeracy were associated with solving the problem in the tablet but not in the model condition. This pattern suggests that higher numeracy is helpful is relaxing the arithmetic constraint in the tablet condition. In the model condition, that constraint is more easily discarded through the manipulation of sets, and hence numeracy contributes to a lesser extent to problem solving success.

Beyond individual differences in creativity and numeracy which may underpin performance in the interactive context, participants may vary in their willingness to interact with the artefacts presented, that is, vary along a dimension of instrumentality. In other research (Vallée-Tourangeau, Abadie, & Vallée-Tourangeau, 2015) we have found evidence that not all individuals react in the same way when presented with manipulable objects. When presented with a pack of cards to support their solving of a complex statistical reasoning problem, some participants only made minimal actions on the cards (i.e., making mainly “marking actions” that had no obvious epistemic or perceptual impact on the physical layout). By contrast, other participants were more likely to engage in transformative actions, that is presentation change actions that significantly changed the perceptual layout. These participants were also more likely to correctly solve the problems they were working on. Future research may investigate differences among people in their willingness or ability to engage with material artefacts.

4.General Discussion

Theoretical accounts of insight problem solving focus on the mental processes that enable participants to overcome an impasse (e.g., Fleck & Weisberg, 2013). A fundamental assumption of these accounts is that problem solving proceeds on the basis of a mental representation of the problem and changes to this representation that may eventuate in an action plan that provides a solution to the problem. An important correlate of this framework is that “representational change processes do not correspond to any particular overt behaviors” (Ohlsson, 2011, p. 113). Since representations must be kept active in some mental workspace, working memory capacity has been implicated in insight problem solving, and indeed positive correlations between measures of working memory and insight problem solving performance have been reported (e.g., Gilhooly & Fioratou, 2009). However, research on insight problem solving commonly employs procedures that prevents or limits interaction with a physical presentation of a problem, a form of cognitive sequestering (Vallée-Tourangeau & Vallée-Tourangeau, 2014).

By limiting or eliminating interactivity with a physical model of the problem, problem solving must inevitably proceed on the basis of a representation of the problem, and hence research efforts complacently validate rather than test the representational assumptions of the prevalent paradigm.

A reviewer of a previous version of this manuscript pointed out that earlier work on insight problem solving involved actions with material artefacts, making specific reference to Duncker's (1945, p. 86) "box problem" and Maier's (1931) "tying strings" problem. Indeed, some of the pioneering work on the psychology of insight problem solving in the first few decades of the 20th Century involved participants (and in the case of Köhler's ethnographic research, chimpanzees) working on problems that were presented with various objects and involved physical interaction. However, this work was conducted from a Gestalt perspective, and neither Duncker or Maier put any emphasis on the constitutive role of interactivity in the genesis of insight. Duncker used the box problem to investigate functional fixedness and compared solution rates between conditions involving pre-utilization (boxes filled with other objects, such as matches) or without pre-utilization (empty boxes). Duncker's analysis did not focus on actions leading to solution (how to affix a candle onto a vertical surface) and his perspective on problem solving was firmly internalist, labouring as sharp a contrast as possible with behaviourist and associative models of problem solving. Maier (1931), like Duncker, casts problem solving as reflecting the re-organisation of mental elements. At the end of his original paper on the two-string problem he writes (p. 193): "The perception of the solution of a problem is like the perceiving of a hidden figure in a puzzle-picture. In both cases, (a) the perception is sudden;

(b) there is no conscious intermediate stage; and (c) the relationships of the elements in the final perceptions are different from those which preceded, i.e., changes in meaning are involved". Maier (1930, 1931) was concerned with how certain hints would spark such a sudden re-organisation, *not* how an agent's action dynamically modified a problem presentation in leading to insight.

Outside the psychologist's laboratory, however, problem solving involves interacting with resources external to the reasoner (Kirsh, 2013); solutions are distilled on the basis of a great deal of interaction with the material world. From an ecological perspective, cognition evolved to ensure the dynamic coupling between an organism and its environment through interactivity (Anderson, 2014; Järvilehto, 1998). We would argue that an important, albeit implicit, reason for failing to comment on the role of interactivity in problem solving is the methodological and theoretical commitment to formulating an explanation in terms of mental processes that transform a mental representation of the world. Solving problems, however, outside the cognitive psychologist's laboratory first involves changing the world. That is, problem solutions are reflected in changes in the world; these physical changes are the evidence of a solution. Problem solving in the world involves tools, maps, models—some readymade models as those used in teaching organic chemistry (Toon, 2011), some reflecting constructions using artefacts at hands, like the table top model of a city's landmarks described in Noë (2012)—and unfolds within a set of spatio-temporal coordinates. Solving problems in the world primarily involves action: To solve problems is to act in the world.

At one level of analysis, participants in the experiments reported here solved the same problem in both the tablet and model conditions. But the cognitive ecologies were very different, and arguably different in the model condition of Experiment 1 and Experiment 2. The cognitive ecosystems (cf. Huchins, 2010) implemented in the tablet and model conditions were populated with different tools, arrayed in a different physical space, prompted and cued a different range of actions. Participants enacted different hunches and explored different paths to solution in these different ecologies. The problem was more easily restructured when participants engaged in model building activity. The genesis of insight can be understood as an enacted phenomenon produced through the interactivity that couples an agent to the material world. A mentalist perspective focusing on internal processes that restructure a mental representation does not alert researchers to the importance of interactivity and the materiality of the artefacts that populate the ecosystem. While a mentalist perspective may acknowledge the role of the environment in shaping internal representations, the associated ontological and methodological commitments make it difficult to predict how problem-solving performance may differ in environments that support different types of interactivity.

4.1 Trajectory to Solution: Singular and Contingent

The idea explored in this paper is that interactivity is an ontological substrate (Steffensen, 2013), that is that new ideas emerge through material engagement. In other words, engaging with the material world is an enabling condition for conceptual change (see also Malafouris, 2015, on plasticity through interaction with artefacts). In the model condition, a hunch is reified into a working hypothesis through interacting with a dynamic problem presentation, using material that encourages the construction of the key item in the problem solution, namely the pens. However, insight is not the mechanical result of certain actions or triggered once the physical presentation of the problem has been configured in a certain way. The importance of an action or of a perceptual cue is granted by its contingent occurrence at a certain point in time; certain actions are important or cues significant by virtue of their temporal context, that is by virtue of what has gone on before. The path to discovery for each solver is a singular and contingent trajectory: no single action or perceptual cue is necessary to break the initial arithmetic impasse.

Given the singularity of the solution path woven through action, a detailed qualitative analysis of each participant’s effort is required, which is beyond the scope of this paper. However single micro-detailed case analyses of individual participants offer telling a telling window onto the origin of new ideas, what Steffensen (2015) termed the *probatonic* principle. That principle is at the heart of Steffensen’s (2013) cognitive event analysis, which we applied to one of the 6 successful participants in Experiment 1 of the present paper; this case study is reported in Steffensen, Vallée-Tourangeau, and Vallée-Tourangeau (2016). Steffensen et al. offered a micro analysis of the actions and evolving physical configurations of the 17A problem: the 10-minute problem solving session was coded in terms of 1291 separate annotations, for an annotation density of 2.152 annotations per second (for comparison, Anzai and Simon’s [1979] “microscopic account“ of a participant solving the Tower of Hanoi was based on 232 annotations during a 90-minute session and thus yielding an annotation density of 0.043 annotations per second). As with all other participants in the model condition, this participant spent the first few minutes building pens (in her case, 1:52 minutes to be exact). In this preparation phase, three overlapping configurations were accidentally produced and were promptly disassembled to restore a configuration involving four non-overlapping enclosures. For the next 3:47 minutes of the problem solving session, the participant attempted to distribute the 17 animals in the four separate enclosures. There followed a 26-second phase during which she took all the zebras and put them in a heap in the middle of the work table. In effect, there was a figure-ground inversion, where the pens rather than the animals became the focus of her attention. She then re-adjusted one of the pens, apparently to fashion it into a less oblong and more circular shape. Because the pipe cleaning material is so light their chenille texture can turn into velcro when coming into contact with other pipe cleaning pieces, and in the process of adjusting one pen, it encroached on an adjacent one and dragged it across to create two overlapping pens. Micro-coding of her action revealed that she proceeded to remove the overlap but inhibited her action just as her hand was about to reach one of the pens. Within the next few seconds, she created two additional overlaps such that the four pens were arrayed in a U shape, each one overlapping with the next, creating three intersections. For the next 1:24 minutes the participant worked out the distribution of the animals such as to meet the problem requirement.

This micro-analysis of one participant working on the 17A problem revealed a few important points. First, no specific action or perceptual configuration taken in isolation leads to a solution. Rather, their role in shaping problem solving efforts depends on the preceding contingent trajectory. For example, overlapping configurations can be encountered but this perceptual information may not trigger a different way of dealing with the problem. Thus, to understand the origin of new ideas and the overcoming of impasse, the temporal trajectory of the participant's progress should be analysed. Second, materiality matters. Thus the properties of the pipe cleaning pieces in themselves guide the problem solver along a certain path, what Ingold (2010) refers to as the textility of making. Third, material engagement creates chance configurations that may prompt different ways of tackling the problem (Kirsh, 2014).

The work reported here does not proceed from a radical enactivist view that aims to jettison all talk of mental representations (e.g., Chemero, 2013; Hutto, 2005). Problem solving reflects a relation between external resources and internal representations; to adapt Malafouris (2015), we would argue that insight is an 'in-between' process. It may well be easier to identify and measure features of the thinking agents rather than the features of the broader cognitive system within which they are embedded. We think, however, that it is not fruitful to attempt an orthogonal segmentation of the different internal and external components of the system that is configured by the meshing of internal and external resources through interactivity. The ontology of insight is relational, internal and external resources are mutually constitutive in a transactional process. Insight is operationalised in terms of agent-induced changes in the external resources.

In a classic paper, Zhang and Norman (1994) discussed the role of external representations in non-insight transformation problem solving. Using a size-reversed Tower of Hanoi task, Zhang and Norman elegantly varied how move selection constraints could be embedded in the physical artefacts employed to represent different isomorphic versions of the task. As the number of rules represented externally increased, problem solving performance—as gauged in terms of solution latencies, moves and error—improved. According to Zhang and Norman external representations expand memory capacity, can be directly perceived, and transform the nature of the task. We certainly concur. However, Zhang and Norman worked with a non-insight problem with a fully specifiable problem space that can explicitly map all the possible moves from the initial to the goal state (which then permits “the explicit decomposition of the representational system of a distributed cognitive task into its internal and external components”, p. 120, that they prescribe). That kind of analysis is not possible with an insight problem. More important, the data presented here point to the importance of interactivity and underscore the contingent nature of the spatio-temporal trajectory of insight problem solving, a trajectory enacted on the basis of interactivity (Vallée-Tourangeau & Vallée-Tourangeau, 2014).

In both experiments, participants were randomly allocated to either condition. As a result, cognitive capacities and thinking dispositions were equated. In Experiment 1, no participant formulated a solution in the tablet condition; their failure to do so cannot be attributed to sub-normal working memory capacity or poor thinking dispositions. Nor could working memory capacity and thinking dispositions explain differences between solvers and non-solvers in the tablet condition. This is likely due to the contingent nature of the problem solving trajectory. As the problem presentation is modified, it dynamically unveils different paths, some promising, some leading to a cul-de-sac. For example, in Experiment 1, some participants in the model condition wove pipe cleaners to create tagliatelle nests, presumably to prevent animals from ‘jumping’ over the pen’s perimeter. Such structures however are not conducive to the creation of overlapping sets. As a result, once participants make this design decision, the problem solving trajectory cannot eventuate in a solution to the problem. Model building in Experiment 2 used pre-formed hoops and as such participants could not build enclosures that prevented the development of overlapping configurations. In addition, we conjectured that hoops might guide reasoners along a more productive trajectory because they are artefacts commonly associated with overlapping configurations. The latter feature in itself, though, was insufficient to spark insight since fewer than 20% of the participants in the tablet condition solved the problem. Rather, it is the physical interaction with the physical elements of the model that is an important driver of insight.

4.2 Methodological Implications and Recommendations

Our findings have important implications for the psychology of problem solving. Understanding how people solve problems must proceed from an appreciation of the dynamic coupling between a reasoner and her environment. The recent integrated framework of insight problem solving offered by Fleck and Weisberg (2013) was informed by verbal protocols of participants as they worked through a series of insight problems, some presented with artefacts that could be manipulated to alter the problem configuration (e.g., the triangle of coins), others not—e.g., the water lilies problem also used in the cognitive reflection test (Frederick, 2005). Fleck and

Weisberg reported but did not comment on the fact that, in their data, restructuring was more likely to occur when participants could manipulate the physical elements of a problem (Vallée-Tourangeau, 2014). They probably did not comment on this feature of their findings because they did not assume it would make a difference: If cognition only takes place in the head, it should not. But the problem solving data reported here clearly show that it makes all the difference. To understand how people solve problems we need to understand how different task ecologies, with different properties and affordances, may encourage different problem-solving trajectories. Qualitative analyses reflecting a detailed coding of actions and the resulting dynamic configuration of the problem presentation—with its shifting topography of affordances—will likely offer a better explanation of how, why and when someone achieves insight (Steffensen et al., 2016). Methodologically this program of research can only proceed by ensuring that interaction with the physical constituents of a problem presentation is possible.

We can anticipate how such interactive task procedures may be integrated and complemented with verbal protocols, to determine how participants' on going narrative anticipate or follow certain actions (and provide important data concerning the phenomenology of insight in interactive environments). In addition, eye-tracking data may offer a better gauge of the allocation of attention (e.g., how shifts in visual attention precede certain actions, how new problem configurations garner sustained attention at different stages of the problem solving trajectory). Such research efforts would then measure action, visual attention, and verbal narrative, considerably enhancing the richness and granularity of the analysis and help identify which actions are intentional and which reflect the relatively unmediated reaction to an environmental affordance. Such method may also reveal the relative proportions of unmediated actions reflecting swift perception-action cycles and those that reflect the intentional implementation of a plan at different stages of the spatio-temporal trajectory.

Acknowledgements

We thank Erica Mundhal for her help in developing some of the experimental material, and Emily Dewar, Pani Fanai-Danesh and Angie Makri with help recruiting and running participants for Experiment 2. We thank the Georgia Tech Attention & Working Memory Lab for the use of the working memory capacity tests employed in Experiment 1.

References

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive Brain*. Cambridge, MA: MIT Press.

Anzai, Y., Simon, H. A. (1979). The theory of learning by doing. *Psychological Review, 86*, 124-140. http://dx.doi.org/10.1037/0033-295X.86.2.124

Batey, M. D. (2007). *A psychometric investigation of everyday creativity*. Unpublished PhD dissertation. University of London.

Baron, J. (1985). *Rationality and intelligence*. New York: Cambridge University Press.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48,* 306-307. http://dx.doi.org/10.1207/s15327752jpa4803_13

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional difference in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 197-253. http://dx.doi.org/10.1037/0033-2909.119.2.197

Carlson, R. A., Avraamides, M. N., Cary, M., & Strasberg, S. (2007). What do the hands externalize in simple arithmetic? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 747–756. http://dx.doi.org/10.1037/0278-7393.33.4.747

Chemero, A. (2013). Radical embodied cognitive science. *Review of General Psychology, 17,* 145-150. http://dx.doi.org/10.1037/a0032923

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. Unpublished PhD dissertation, McGill University.

Davidson, J. (1995). The suddenness of insight. In R. J. Sternberg and J. E. Davidson (Eds.), *The nature of insight* (pp. 125-155). Cambridge MA: The MIT Press.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58(5), i-113. http://dx.doi.org/10.1037/h0093599

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309-331. http://dx.doi.org/10.1037/0096-3445.128.3.309

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the Subjective Numeracy Scale. *Medical Decision Making, 27*, 672-680. http://dx.doi.org/10.1177/0272989X07304449

Fleck, J. I., & Weisberg, R. W. (2004). The use of verbal protocols as data: An analysis f insight in the Candle problem. *Memory and Cognition*, *32*, 990-1006. http://dx.doi.org/10.3758/BF03196876

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

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25-42. http://dx.doi.org/10.1257/089533005775196732

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. http://dx.doi.org/10.1080/02724980543000051

Gilhooly, K. J., & Fioratou, E. (2009). Executive functions in insight versus non-insight problem solving: An individual differences approach. *Thinking and Reasoning*, *15,* 355-376. http://dx.doi.org/10.1080/13546780903178615

Guilford, J.P., Christensen, P.R., Merrifield, P.R., & Wilson, R.C. (1978). *Alternate uses: Manual of instructions and interpretations*. Orange, CA: Sheridan Psychological Services.

Guthrie, L. G., Vallée-Tourangeau, F., Vallée-Tourangeau, G., & Howard, C. (2015). Learning and interactivity in a transformation problem*. Memory and Cognition, 43,* 723-735. http://dx.doi.org/10.3758/s13421-015-0504-8

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making, 8*, 188-201.

Hunt, T. E., Clark-Carter, D., & Sheffield, D. (2011). The development and part validation of a UK scale for mathematics anxiety. *Journal of Psychoeducational Assessment, 29*, 455-466. http://dx.doi.org/10.1177/0734282910392892

Hutchins, E. (2010). Cognitive ecology. *Topics in Cognitive Science, 2*, 705-715. http://dx.doi.org/10.1111/j.1756-8765.2010.01089.x

Hutto, D. D. (2005). Knowing what? Radical versus conservative enactivism. *Phenomenology and the Cognitive Sciences, 4*, 389-405. http://dx.doi.org/10.1007/s11097-005-9001-z

Ingold, T. (2010). The textility of making. Cambridge Journal of Economics. 34, 91-102. http://dx.doi.org/10.1093/cje/bep042

Järvilehto, T. (1998). The theory of the organism-environment system: I. Description of the theory*. Integrative Physiological and Behavioral Science, 33,* 321-334. http://dx.doi.org/10.1007/BF02688700

Kaplan, C.A., & Simon, H.A. (1990). In search of insight. *Cognitive Psychology, 22*, 374-419. http://dx.doi.org/10.1016/0010-0285(90)90008-R

Kirsh, D. (2013). Thinking with external representations. In S. J. Cowley, & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: Interactivity, cognition and human artifice* (pp. 171-194). Dordrecht: Springer.

Kirsh, D. (2014). The importance of chance and interactivity in creativity. Pragmatics & Cognition, 22, 5-26. http://dx.doi.org/10.1075/pc.22.1.01kir

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology*: *Learning, Memory and Cognition*, *25*, 1534-1555. http://dx.doi.org/10.1037/0278-7393.25.6.1534

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

Köhler, W. (1925/1957). *The mentality of apes.* London: Penguin Books.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. New York: Cambridge University Press.

Lydersen, S., Fagerland, M. W., & Laake, P. (2009). Recommended tests for association in 2×2 tables. *Statistics in Medicine, 28*, 1159-1175. http://dx.doi.org/10.1002/sim.3531

Maier, N. R. F. (1930). Reasoning in humans. I. On Direction. *Journal of Comparative Psychology*, *10*, 115-143. http://dx.doi.org/10.1037/h0073232

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology, 12*, 181-194. http://dx.doi.org/10.1037/h0071361

Manktelow, K. I., & Over, D. E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition*, *39*, 85-105. http://dx.doi.org/10.1016/0010-0277(91)90039-7

Malafouris, L. (2013). *How things shape the mind: A theory of material engagement*. London: The MIT Press.

Malafouris, L. (2015). Metaplasticity and the primacy of material engagement. *Time and Mind: The Journal of Archaeology, Consciousness and Culture, 8*, 351-371. http://dx.doi.org/10.1080/1751696X.2015.1111564

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, *15*, 238-246. http://dx.doi.org/10.3758/BF03197722

Noë, A. (2012). *Varieties of presence*. Cambridge, MA: Harvard University Press.

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working memory and intelligence-their correlation and their relation: Comment on

Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 61-65. http://dx.doi.org/10.1037/0033-2909.131.1.61

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. T. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (pp. 1- 44). Hemel Hempstead, UK: Harvester Wheatsheaf.

Ohlsson, S. (2011). *Deep learning: How the mind overrides experience*. New York: Cambridge University Press.

Patton, J.H., Stanford, M.S., & Barratt, E.S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology, 51*, 768-774. http://dx.doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Sirota, M., Juanchich, M., & Hagmayer, Y. (2014). Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychonomic Bulletin & Review, 21*, 198-204. http://dx.doi.org/10.3758/s13423-013-0464-6

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161-188. http://dx.doi.org/10.1037/0096-3445.127.2.161

Steffensen, S. V. (2013). Human interactivity: Problem-solving, solution probing and verbal patterns in the wild. In S. J. Cowley, & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: Computation, interactivity and human artifice* (pp. 195-221). London: Springer-Verlag.

Steffensen, S. V. (2015). Cognitive probatonics: Towards an ecological psychology of cognitive particulars. *New Ideas in Psychology*, forthcoming. http://dx.doi.org/10.1016/j.newideapsych.2015.07.003

Steffensen, S. V., Vallée-Tourangeau, F., & Vallée-Tourangeau, G. (2016). Cognitive events in a problem-solving task: Qualitative methods for investigating interactivity in the 17 animals problem. *Journal of Cognitive Psychology*, *28*, 79-105. http://dx.doi.org/10.1080/20445911.2015.1095193

Toon, A. (2011). Playing with molecules. *Studies in History and Philosophy of Science*, *42*, 580-589. http://dx.doi.org/10.1016/j.shpsa.2011.08.002

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498-505. http://dx.doi.org/10.3758/BF03192720

Vallée-Tourangeau, F. (2012). Utilities in the 2-4-6 task. *Experimental Psychology*, *59*, 265-271. http://dx.doi.org/10.1027/1618-3169/a000152

Vallée-Tourangeau, F. (2013). Interactivity, efficiency, and individual differences in mental arithmetic. *Experimental Psychology*, *60*, 302-311. http://dx.doi.org/10.1027/1618-3169/a000200

Vallée-Tourangeau, F. (2014). Insight, materiality and interactivity. *Pragmatics & Cognition, 22,* 27-44. http://dx.doi.org/10.1075/pc.22.1.02val

Vallée-Tourangeau, G., Abadie, M., & Vallée-Tourangeau, F. (2015). Interactivity fosters Bayesian reasoning without instruction. *Journal of Experimental Psychology: General, 144,* 581-603. http://dx.doi.org/10.1037/a0039161

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

Vallée-Tourangeau, G., & Vallée-Tourangeau, F. (2014). The spatio-temporal dynamics of systemic thinking. *Cybernetics and Human Knowing, 21*, 113-127.

Vallée-Tourangeau, F., & Vallée-Tourangeau, G. (2016). Thinking and deciding beyond the brain. In N. Galbraith, E. Lucas, & D. Over (Eds.), *The thinking mind: The use of thinking in everyday life (essays in honour of Ken Manktelow*). Hove: Psychology Press.

Weisberg, R. W. (1995). Prolegomena to theories of insight in problem solving: A taxonomy of problems. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight* (pp. 157-196). Cambridge MA: MIT Press.

Weisberg, R.W. (2015). Toward an integrated theory of insight in problem solving. *Thinking & Reasoning, 22*, 5–39. http://dx.doi.org/10.1080/13546783.2014.886625

Weller, A., Villejoubert, G., Vallée-Tourangeau, F. (2011). Interactive insight problem solving. *Thinking & Reasoning*, *17*, 429-439. http://dx.doi.org/10.1080/13546783.2011.629081

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, *18*, 87-122. http://dx.doi.org/10.1207/s15516709cog1801_3