Insight as Discovery

Frédéric Vallée-Tourangeau, Wendy Ross, Renata Ruffatto Rech,

Department of Psychology

Kingston University

and

Gaëlle Vallée-Tourangeau

Department of Management

Kingston University

Author Note

Abstract

Performance on the triangle of coins problem was observed when participants could physically change the configuration of the problem as they explored potential solutions or when the problem remained fixed and new configurations could only be simulated rather than enacted. Participants were filmed as they worked on the triangle of coins problem with a bespoke computer interface for a ten-minute period. One group of participants could touch and move the 'coins' on the screen, thereby enacting changes to the physical configuration of the problem (high interactivity) while a second group could not (low interactivity). In Experiment 1, participants could record as many possible solutions as they wished and receive feedback for incorrect ones, whereas in Experiment 2, they were limited to only one announcement; if the proposed answer was incorrect, the session was terminated. In both experiments, solution rates were generally better in the high than in the low interactivity environment, although participants were slower and made more moves before announcing their solution in the high interactivity condition of Experiment 2. Detailed analysis of the video data for the participants in the high interactivity task environment enabled the granular mapping of the changes to the physical configuration of the problem, unveiling a gradual appreciation of the solution, punctuated by many discontinuities in move latencies.

Insight as Discovery

In an early section of *The mentality of apes*, titled 'The use of implements', Köhler describes the box test: A banana is nailed to the roof, too high for the animals to reach, and a box with an open side is placed in the middle of the room. Sultan, famously, made a quick meal out of this situation: "from the momentary pause before the box to the first bite into the banana, only a few seconds elapsed, a perfectly continuous action after the first hesitation (p. 40)". It is observations such as these that encouraged psychological researchers to study insight as the sudden realisation of a solution and its fluid implementation. Its iconic representation, a lightbulb suddenly flashing brightly over what was hitherto a darkness of frustration and incomprehension, seems to aptly symbolize the phenomenon[1].

However, on the very next page, another ape, Koko, is the subject of Köhler's observations: Koko's performance on the box test is described by Köhler as 'most curious' (p. 41). Koko's problem-solving process is altogether different; the solution appears to have taken shape very slowly, full of false starts, hazy hunches, and tentative movements. Köhler writes: "Koko showed a lively interest in the fruit, but none—at first—in the box, and when he began to move it, he did not appear apathetic but uncertain: there is only one (colloquial) expression that really fits his behaviour at that juncture: 'it's beginning to dawn on him!'" (p. 42). For Koko, then, the lightbulb did not shine quickly or very brightly: How different might have research on problem solving developed had Köhler made Koko the star of his book. The key point is that insight is not always achieved through a mental leap, it can also manifest itself at the end of a non-linear exploration of both the conceptual and the material problem space.

The Explorer Metaphor

---

[1] Take the instructions to participants in Danek et al. (2014, p. 4): "As an example, imagine a light bulb that is switched on all at once in contrast to slowly dimming it up."

To illustrate further the importance of considering insight as a process of situated discovery, let's exchange the sudden flash of light metaphor with another one, that of the explorer, and in the process replace the lightbulb with a different symbol, perhaps a hammer and pick, in a classic heraldic cross. Our explorer is fully equipped, perhaps on an archaeological or paleontological mission, certainly not one directed from her armchair. The goal of her exploration may be sharply defined or poorly specified or completely misconceived (and the nature of the goal may itself shift during her mission). The landscape she travels constrains and guides her path, some terrain inviting exploration, other too treacherous to cross. On her journey she may see unexpected things; at times her movements reflect well-rehearsed methods, at others she engages in unplanned actions. Her efforts can unearth promising leads, and lucky breaks may encourage her to dig a certain spot. She may stumble upon something completely unexpected (Austin's [1979] Kettering principle), and along the way discover something interesting perhaps even important, but not something she had set out to find.

To do justice to the process of discovery, we should document her physical toils, the tools employed, what she saw, the serendipitous moments and tortuous progress that paved the stages of her discovery; these cannot be predicted a priori by assessing her intelligence or her working memory capacity before she sets off on her journey. Her concept of what she is looking for remains underspecified until she makes her discovery; her anticipation of what she is looking for becomes sharper as she makes it. Epistemological and ontological considerations are not clearly demarcated: The physical manipulation and perceptual appraisal of the object once discovered sharpen her own mental representation of it, which in turn shapes her analysis of what she sees, reinforcing the reality of the object of discovery, in a tight cycle of reciprocity.

The contrast between an armchair explorer and a field one invites a distinction between first and second order problem solving. The latter is based on mental manipulation of abstract representations of the world; the exploration is mentally simulated based on already obtained knowledge. It is second order because it does not proceed from actual engagement with the world but rather from a simulated one. First order problem solving in turn, proceeds with and through a physical world: it is thinking with things, or 'thinging' as Malafouris (2020) calls it. Exploration of the problem space is not virtual or theoretical but rather enacted physically. These actions can change the physical presentation of the problem, offering new feedback that cues new ways of acting on the problem. The nature of the processes underlying second order problem solving are inescapably mental, internal to the reasoner (or explorer) because she is cut off from engagement with the world. The processes that undergird first order problem solving are inescapably systemic: The cognitive resources involved in first order activities are distributed and emergent. The dynamic engagement with the world modifies the representational and storage burden on the explorer—since the world is there to see, it need not be represented—and interim proto solutions are intermediate way stations that guide the next action (proto solutions as 'boundary objects', after Fiore & Wiltshire, 2016). First order problem solving can proceed along different time scales (e.g., moving fast, thinking slow), is transformative, proceeding from cycles of doing and undergoing (after Dewey, 1934; cited in Glăveanu et al., 2013), and can exploit felicitous but serendipitous juxtapositions of problem elements (for additional characteristics of first order problem solving see Vallée-Tourangeau & Vallée-Tourangeau, 2020a, Vallée-Tourangeau & March, 2019).

In first order problem solving, thinking and actions are interwoven, and both take shape with and through a dynamic physical environment. The explorer metaphor suggests that the journey to discovery, that is the spatio-temporal path that led a participant to discover a solution, will be singular and contingent. While problem solving performance can be

captured with aggregate measures such as solution rates and latencies to discovery, the nature

of the trajectory in terms of actions made, areas explored, and changes to the environment as

a result of those actions invites a more qualitative narrative. First order problem solving may

thus be amenable to a mix of quantitative and qualitative analyses. For example, Ross and

Vallée-Tourangeau (2020) employed such a mix to profile the complex, contingent and

dynamic nature of behaviour as participants engage in a Scrabble-like word production task

with moveable letter tiles. Thus, to study first order problem solving in a laboratory

environment, it is important to design problem solving tasks that afford the interaction with a

physical and malleable problem presentation. First order problem solving tasks can be

instrumentalized to provide data that reveals the physical traces of problem solving, to map

the genesis of a new idea (Vallée-Tourangeau & Vallée-Tourangeau, 2020b). Failing to

record and code with some precision how and to what degree participants interact with the

physical model of the problem will not reveal the interim proto solutions that are reified

physically; the origin of a new idea, in this instance the solution to a problem, will remain

inaccessible (for an example, see Chuderski, Jastrzębski, & Kucwaj, 2020).

The distinction between first and second order problem solving is not clearly or

explicitly made in current problem-solving research. A particularly informative example is

found in Fleck and Weisberg's (2013) important integrative framework paper. These

researchers collected verbal protocols from participants working on a set of five classic

insight[2] problems, namely the socks, the triangle of coins, the trees, the cheap necklace, and

the lilies. The first and last problem were presented as a verbal description only, while the

middle three were presented with "physical props" (see Fleck & Weisberg, 2013, pp. 446-

447). This difference in procedure, the mix of first and second order problem solving, is not

---

[2] This term is used here as a shorthand for the traditional problems employed in insight problem-solving
research; this does not mean that the solution of these problems involves a classic insight sequence with a
concomitant 'aha' subjective experience.

explicitly discussed by these researchers; the two verbal protocols cited at length in their

paper involved participants manipulating artifacts and observing the resulting change to the

physical presentation of the problem (see p. 452 for the verbal protocol of a participant

working on the triangle of coins problem, accompanied by a 6-panel figure illustrating the

changes in the configuration of the problem; p. 454 for the one of a participant working on

the trees problem). Fleck and Weisberg's data on the rate of switching heuristics and the

different types of restructuring and indeed the paucity of the classic insight sequence (clearly

reported in their Table 3 on p. 451) are not presented as a function of the interactivity

afforded by the problem solving environment. Yet, it is interesting to note that the problem

that elicited the most switching heuristics is the trees (which proceeded from the

manipulation of artifacts) and that data-driven restructuring was only observed with problems

that afforded interactivity (see Vallée-Tourangeau, 2014, for a detailed reanalysis of the data

reported by Fleck & Weisberg, 2013).

The Triangle of Coins

In this problem 10 coins are arranged in a triangular shape pointing down (see Figure 1,

top panel, for an illustration of the problem as configured in the experiments reported here):

Participants must identify which three coins can be moved to rotate the orientation of the

triangle such that it points up rather than down. To solve the problem, coins from the central

hexagonal rosette should *not* be moved; the solution involves moving the corner coins. Fleck

and Weisberg gave participants artifacts that corresponded to the 10 coins to move in a

manner of their choosing to help them discover the solution: The overall solution rate was

69% over a 10-minute period (see also Chronicle, MacGregor, & Ormerod, 2004).

Vallée-Tourangeau (2017) sought to offer a sharper contrast between a first and second

order problem-solving task environment by manipulating the level of interactivity. In a low

interactivity condition, the problem was presented on a side of A4 which displayed an image

of 10 coins in a south-pointing triangular formation. Each coin was located in a cell of a 9x9

chequer board, with numbered rows (1-9) and columns labelled with letters (A-I); the

pictured coins were also labelled with individual letters. While keeping their hands flat in

front of them on the work surface, participants were given five minutes to announce the

coordinates to which three coins could be moved to create a new triangular shape pointing

up. The overall solution rate in the low interactivity condition was 23%. In the high

interactivity condition, the chequer board was reproduced on a laminated side of A3 and

actual wooden tokens adorned with letters were used to create the triangular shape.

Participants were invited to move the tokens as they saw fit to help then discover the

coordinates of the three coins that would change the orientation of the triangle. The overall

solution rate in the high interactivity condition was 60%. Thus, as a second order problem

solving exercise the triangle of coins appeared particularly challenging; solution rate was

nearly tripled in a first order problem-solving environment. However, there were a number of

inelegant procedural details in this experiment. For one, the problem was presented on a

much larger area in the high than in the low interactivity condition (A3 vs. A4, respectively).

Second, participants in the high interactivity condition experienced seeing the coins in

different configurations on the board, perhaps facilitating their projection to different

locations; participants in the low interactivity condition could not have seen coins distributed

elsewhere on the board. Finally, the allocated time to discover the solution was five minutes,

half the time allocated to participants in Fleck and Weisberg.

        The Present Experiments. The two experiments reported here sought to redress these

procedural limitations. In both experiments, participants worked on the triangle of coins

problem for a ten-minute period. The problem was presented with a bespoke computer

program operating on a touch screen interface. Participants in a high interactivity condition

could move the tokens on screen to discover the solution (by touching and dragging them

with their finger), while those assigned to a low interactivity condition could not. Prior to working on the triangle of coins, participants in both conditions completed a few warmup games dragging the coins on the screen to predetermined positions, although none of the resulting shapes were triangular. Thus, the same apparatus was used in both interactivity conditions; in addition, all participants were pre-exposed to coins placed in different locations on the electronic surface.

In Experiment 1 participants could record as many incorrect answers as they wished while working on the problem, and receive feedback; as a result, participants in the low interactivity condition overcame to some extent the constraint on interaction by making significantly more incorrect announcements than participants in the high interactivity condition. Feedback was of a binary rather than gradual nature: participants were not told which of the tokens was in the wrong position but simply that the whole configuration was right or wrong. We thus designed, pre-registered (https://aspredicted.org/blind.php?x=si9k5j) and ran Experiment 2, which is a replication of the first experiment, but with the added constraint that participants in both conditions were allowed only one attempt to state their answers during the test session; if that announcement was incorrect, then the session terminated.

We predicted that performance, as measured in terms of solution rate and solution latencies would be substantially influenced by the level of interactivity, such that more participants would solve the triangle of coins with high interactivity and do so faster. While we expected a difference in performance owing to the level of interactivity afforded by the task, little is known about how interactivity may foster insight-qua-discovery of a new idea and whether such an insight is always a sudden, light bulb moment or whether it is the result of a long quasi-progressive journey to the solution. To shed some light on the participants' exploration, we mapped their journey to discovery (and failure) with the detailed coding of

the changes in the physical configuration of the problem. Thus, in addition to quantitative analyses of aggregate indices of performance, we use the material traces in the high interactivity environment to offer more detailed case studies of how participants solved or failed to solve the problem.

Method

Participants

Undergraduate and postgraduate psychology students received course credits for their participation. Sixty-six participants were recruited for Experiment 1; six participants quit during the task, two in the low and four in the high interactivity condition, resulting in samples of 31 participants (24 females) and 29 participants (26 females) in the low and high interactivity condition, respectively (M_{age} = 25.4 , SD = 7.1). Seventy-one participants were recruited for Experiment 2; one participant refused to be filmed (in the low interactivity condition), one participant refused to move the coins (in the high interactivity condition), and the computer tablet malfunctioned for one participant resulting in samples of 35 participants (31 females) and 33 participants (30 females) in the low and high interactivity condition, respectively (M_{age} = 25.3, SD = 8.8).

Materials

We ran the programme on a Window Surface Pro 2 (x-64 -based PC) operating Windows 10 Pro with a 10.6″ touch sensitive screen. The screen was set at an approximately 45° angle and at approximately 45cm from the participant although there were no constraints on the participants' body movements in either condition, so this distance did not remain constant within a session or across participants. The detachable keyboard remained attached throughout the study to facilitate inputting demographic data, but the participants were directed to use the touch screen interface. The programmed application used jQuery and CSS

to manipulate HTML. The high interactivity version used a library called Draggable JS, in

order to drag and drop the tokens.
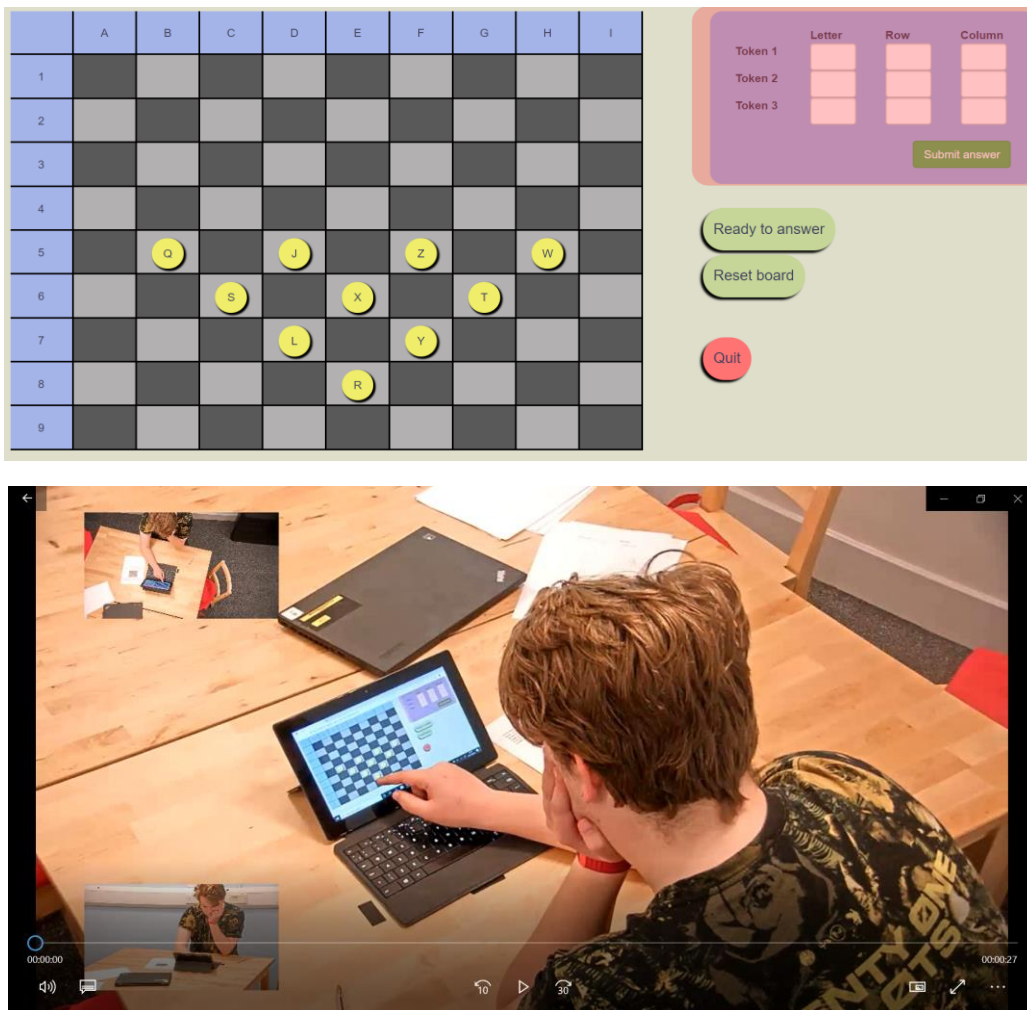
Figure 1. Top panel: Tablet interface on which the triangle of coins was presented (the solution involves moving the three corner coins, for example, R to E4, W to H7 and Q to B7); bottom panel: screen shot from the resulting video from which the changes in the physical configuration of the problem were coded.

Procedure

Participants were tested individually in an observation lab equipped with overhead

cameras. All participants were informed that the test session would be filmed, and all signed

a consent form for the use of the resulting videos in this research and public presentation.

Participants were allocated to either a high or low interactivity condition in a strict alternating

manner. The instructions for the triangle of coins read as follows:

As you can see below, 10 coins are arranged in the shape of an arrow head that points down. The goal is to identify three tokens, and only three, that can be moved to change the triangle such that the arrow shape points up. The shape will remain exactly the same but pointing up rather than down. Each of the ten tokens are labelled with an individual letter. You need to identify which tokens to move as well as their destination on the grid (which has numbered rows and labelled columns).

In the low interactivity condition the instructions continued:

While you are looking at the screen and thinking about possible solutions, it's important to keep your hands flat in front of you. When you think you have the answer, please click on the button 'Submit your answer' and the researcher will enter your guess.

In turn participants allocated to the high interactivity condition read:

You will complete the task on the computer. To help you solve the problem, you can move the tokens to any position on the grid. You will be able to reset the board at any point using the reset button at the side of the screen. When you think you have the answer, please click on the button 'Submit your answer' and the researcher will enter your guess.

In Experiment 1 participants were free to offer as many suggested responses as they wished; in Experiment 2, they were restricted to making one announcement, that is their first suggestion would be taken as their final answer. The participants in both conditions were required to state their answer verbally; the movements in the high interactivity condition were designed to scaffold problem solving and participants were not required to fully solve the problem through moves before announcing their solution. Participants in both conditions sought a solution/worked on the problem with a tablet on which the problem was illustrated (see Figure 1). Before they engaged with the problem, participants in both conditions were invited to interact with the tablet by playing three 'games'. These games were designed to help participants in the high interactivity condition practice touching and moving the coins on the tablet; since this experience might have favoured these participants (through the exploration of different configurations, although none triangular or suggestive of the

solution) participants in the low interactivity condition also played these games before working on the triangle of coins.

In Experiment 2, following an incorrect announcement or at the end of the working session if no announcements had been offered (viz. after ten minutes elapsed), participants in both conditions were given a post task comprehension question. They had to choose, among four possible options, which coin configuration corresponded to the correct solution (the options are illustrated in Figure 2). In turn, if participants in either condition announced a correct answer, they were invited to rate the nature of their experience along four dimensions: happiness "I feel very happy to have solved the problem"), suddenness ("The solution came quite suddenly, as if it jumped directly into my mind"), relief ("When I solved the problem, I experienced a feeling of relief combined with a feeling of happiness after a phase of effort"), and certainty ("I was certain that my solution was correct before making my announcement") on a scale of 1 ("Disagree") to 7 ("Agree").
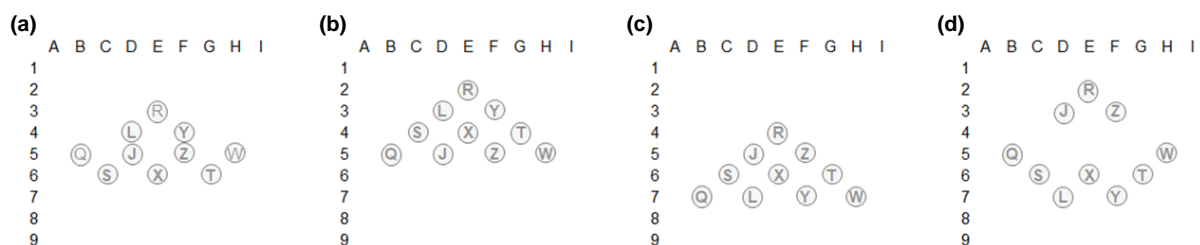
Figure 2. The post task comprehension question in Experiment 2: (c) is the correct answer.

In addition to working on the triangle of coin, participants in both experiments completed the Object-spatial Imagery Questionnaire (OSIQ; Blajenkova, Kozhevnikov, & Motes, 2006) and the Image Control and Recognition Task (ICRT; Irving, Barry, LeBoutillier, & Westley, 2011). The presentation order of these two tasks was counterbalanced in both experiments such that for half of the participants, the test session began with the OSIQ, followed by the triangle of coins, and closed with the ICRT; for the other half the session began with the ICRT, followed by the triangle of coins. and closed with

the OSIQ. The OSIQ is a 30-item self-report measure of visuo-spatial imagery. Participants rate statements such as "I can close my eyes and easily picture a scene that I have experienced" on a scale from 1 ('strongly disagree') to 5 ('strongly agree'). The scale showed good reliability: Cronbach's alpha of .786 and .780 in Experiments 1 and 2, respectively. The ICRT is a guided synthesis task (see Finke, Ward, & Smith, 1992). Participants listen to instructions that describe a shape. For example, the first practice item ('stickman') invites participants to imagine a shape described as follows: "Imagine a plus sign; add a circle to the bottom of the vertical line so it is touching; add a capital 'V' to the top of the vertical line so it is touching; rotate the entire share 180°". Participants are then asked to name the shape and then draw it. The ICRT is composed of 8 items: the task is scored with 1 mark for each item named and 1 mark for each item drawn, and hence scores can range from 0 to 16. The test showed acceptable reliability: Cronbach's alpha of .771 and .685 in Experiments 1 and 2, respectively.

The total session time was approximately 30 minutes.

Results

The solution rates at the 5- and 10-minute mark in Experiment 1 are reported in the top half of Table 1: at the 5-min mark, 19 out of 29 (66%) participants in the high interactivity condition had solved the problem, while 12 out of 31 (39%) participants in the low interactivity condition did so, χ^2 (1, 60) = 4.31, p = .038. Of the participants who solved it under five minutes, four announced the correct solution after three or fewer moves in the high interactivity condition, and two participants solved it under 60 seconds in the low interactivity condition. It is important to note that participants' prior knowledge of the problem was not assessed in either experiment; those who solved it with three or fewer moves might in fact have had some knowledge of the problem and of its solution. At the 10-min mark, solution rates were very similar in the two conditions: 21 (or 68%) and 20 (or

65%) solved the problem in the high and low interactivity condition, respectively, χ^2 (1, 60)

= 0.43, p = .511. In Experiment 1, the average solution latency was faster (M = 143.0s, SD =

127.6s) in the high interactivity group than in the low interactivity group (M = 232.8s, SD =

144.9s), $t(39)$ = 2.11, p = .041 (all tests two-tailed). Over all participants (i.e., including

solvers and non-solvers), more announcements were made in the low (M = 2.15, SD = 1.18)

than in the high interactivity condition (M = 1.19, SD = 0.51), $t(39)$ = 3.40, p = .002.

Table 1

Solution Rates after 5 and 10 Minutes in the Low and High Interactivity Condition in

Experiment 1 and Experiment 2 along with Mean Solution Latencies (and Standard

Deviations).

		Solution Rate						Solution Latency (s)		
		5-min			10-min					
		n	%	χ^2	n	%	χ^2	M	SD	t
Experiment 1	Low	12	39%	4.31 *	20	65%	0.43	232.8	144.9	2.11 *
	High	19	66%		21	68%		143.0	127.6	
Experiment 2	Low	10	29%	3.73	16	46%	3.99 *	217.0	170.1	-0.77
	High	17	52%		23	70%		255.0	136.4	

Note * p < .05

In Experiment 1, successful participants in the low interactivity condition made more

announcements than those in the high condition; feedback from an incorrect announcement is

informative, and hence provided some form of interactivity (and participants in the low

interactivity condition availed themselves of this opportunity more than those in the high

interactivity condition). Experiment 2 eliminated this feature of the procedure of Experiment

1: we predicted that this would widen the performance gap between the two conditions, at

both the 5-min and 10-min mark. We also predicted that solution latencies would be faster in

the high interactivity condition. The solution rates at the 5- and 10-minute mark in

Experiment 2 are reported in the bottom half of Table 1: 10 participants (or 29%) solved the

problem in the low interactivity condition at the 5-min mark, while 17 (or 52%) did so in the

high interactivity condition. While the pattern is consistent with the solution rates observed in

Experiment 1, it was not significant, χ^2 (1, $N = 68$) = 3.73, $p = .053$. Of the participants who

solved it under five minutes, three solved it with three or fewer moves in the high

interactivity conditions, and three participants solved it under 60 seconds in the low

interactivity condition. At the 10-minute mark, there were 16 (46%) solvers in the low

interactivity condition, and 23 (or 70%) solvers in the high interactivity condition, χ^2 (1, $N =$

68) = 3.99, $p = .046$. Thus, our first hypothesis was partially supported. However, our second

hypothesis concerning solution latencies was not supported by the data: the mean solution

latencies did not differ between the low ($M = 217.0$, $SD = 170.1$) and high interactivity

condition ($M = 255.0$, $SD = 133.1$), $t(37)$ =- 0.77, $p = .444$. However, seven participants in

the high interactivity condition in Experiment 2 traced out the problem solution with the

movable tokens more than once before committing to submitting their answer, a careful

manoeuvre considering the one announcement constraint, which may explain the longer

latencies to solution in that condition. The mean number of moves for solvers in the high

interactivity condition in Experiment 2 was 24.7 ($SD = 21.7$) compared to 14.0 ($SD = 13.2$) in

Experiment 1, $t(40) = 1.95$, $p = .059$.

 The number and type of moves made by each participant was coded from the videos.

This enabled us to map the dynamic changes to the problem configuration over time; we will

provide detailed examples of these changes in a later section. We calculated the proportion of

moves involving the corner coins—the solution can only be obtained by moving those

coins--in quartile blocks, that is for the first 25% of the moves by a given participant, the

second 25% and so on. Figure 3 illustrates how the proportion of moves involving the corner

coins gradually increased over time for the successful participants although not for the

participants who did not solve the problem; this pattern was observed in both experiments. A

2 (success) by 4 (quartile) mixed analysis of variance (ANOVA) revealed that in Experiment 1, the main effect of quartile was significant, $F(3, 63) = 3.26$, $p = .037$, $\eta_p^2 = .125$, although not the main effect of success, $F(1, 21) = 3.79$, $p = .065$, $\eta_p^2 = .153$; the interaction was significant, $F(3, 63) = 4.78$, $p = .005$, $\eta_p^2 = .185$. In Experiment 2, the main effect of quartile was significant, $F(3, 75) = 4.59$, $p = .005$, $\eta_p^2 = .155$, as was the main effect of success, $F(1, 25) = 16.97$, $p < .001$, $\eta_p^2 = .404$, and the interaction between quartile and success, $F(3, 75) = 2.759$, $p = .048$, $\eta_p^2 = .099$.
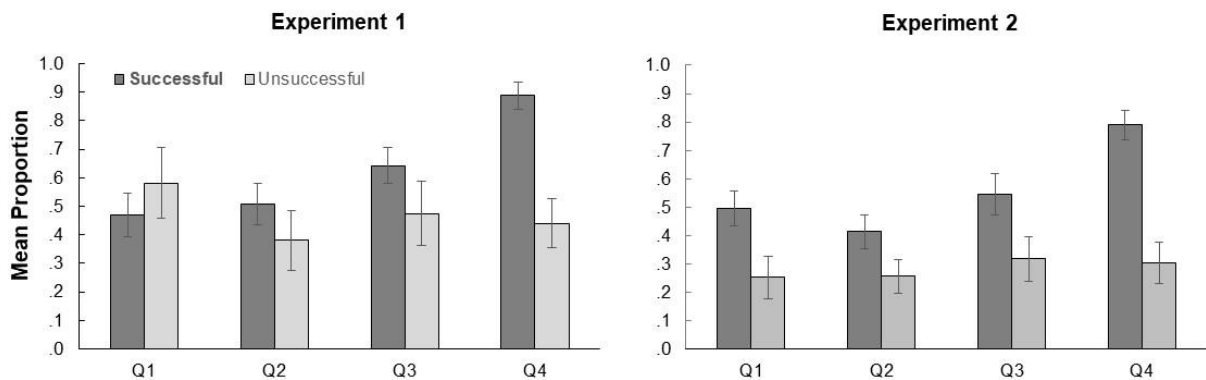
Figure 3. Mean proportion of moves involving the corner coins in each of the four quartiles of moves until a solution is announced for successful and unsuccessful participants. Error bars are standard errors of the mean.

Task Comprehension

The video evidence from Experiment 1 suggested that some participants in the high interactivity condition did not properly understand the goal and/or instructions. (An ancillary benefit of designing first order problem procedures: They can also reveal with some degree of granularity the nature of the unsuccessful participants' confusion.) For example, one participant kept producing shapes, many far from triangular, and making incorrect announcements; this pattern continued for the entire test session. To measure the degree of incomprehension, participants in Experiment 2 who failed to provide a solution to the problem, or who announced an incorrect one, were asked to answer a post-task

comprehension question to determine whether they could recognize the correct answer; participants were presented four coin configurations, and were asked to pick which one provided the solution to the problem. The choice frequencies are shown in Table 2[3]. Among the unsuccessful participants in the low interactivity group, eight (or 44%) chose the configuration that corresponded to the right answer, while those in the high interactivity group, two (or 20%) chose the correct one; however, the choice frequencies did not differ significantly[4] between the two conditions, χ^2 (3, $N = 28$) = 4.85, $p = .183$.

Table 2

Post-task Comprehension Answer Frequencies (and Percentages) among the Unsuccessful Participants in the Low and High Interactivity Group

	A		B		C		D	
Group	n	%	n	%	n	%	n	%
Low	0	0%	9	50%	8	44%	1	6%
High	2	20%	5	50%	2	20%	1	10%

Note: C is the correct answer

Insight Ratings

In Experiment 2, participants who announced a correct answer were asked to rate their experience along four dimensions, (i) happy, (ii) sudden, (iii) relief, and (iv) certain. The mean ratings for each dimension in both interactivity conditions are illustrated in Figure 4. The insight experience as measured in terms of these four dimensions did not appear to differ

---

[3] Note: the post-task comprehension was accidentally not recorded for one unsuccessful participant in the low interactivity condition (P56).

[4] Over 60% of the cells have expected count lower than 5 in this analysis. To correct for this, choice preferences were grouped in terms of correct and incorrect ones (viz. A+B+D vs C). The percentage of cells with expected count lower than 5 was reduced to an acceptable 25%: χ^2 (1, $N = 28$) = 1.67, $p = .196$. Eliminating the participants who could not solve the problem because they could not recognize the correct answers once presented to them from the overall sample, yields an even number of 25 participants in both conditions. The solution rates at the 5-min mark (Low = 40%, High = 68%, χ^2 [1, $N = 50$] = 3.95, $p = .047$) and 10-min mark (Low = 64%, High = 92%, χ^2 [1, $N = 50$] = 5.71, $p = .017$) were significantly greater in the high interactivity condition.

as a function of the level of interactivity; however, the suddenness dimension appeared to have been rated the lowest. A 2 (interactivity level) by 4 (insight dimensions) mixed ANOVA confirmed these impressions: the main effect of group was not significant, $F < 1$, but the main effect of dimension was, $F(3, 111) = 8.53$, $p < .001$, $\eta_p^2 = .187$; the interaction was not significant, $F < 1$. Post hoc tests (with a Bonferroni correction) confirmed that ratings along the sudden dimension were significantly lower than ratings along all other dimensions (least significant $p = .046$, sudden vs. relief), while ratings among all other three dimensions did not differ significantly (smallest non-significant $p = .059$, happy vs. relief). It is worth noting that our instructions to participants characterised the dimension of relief as including a feeling of happiness which may make the difference between these two dimensions harder to interpret.
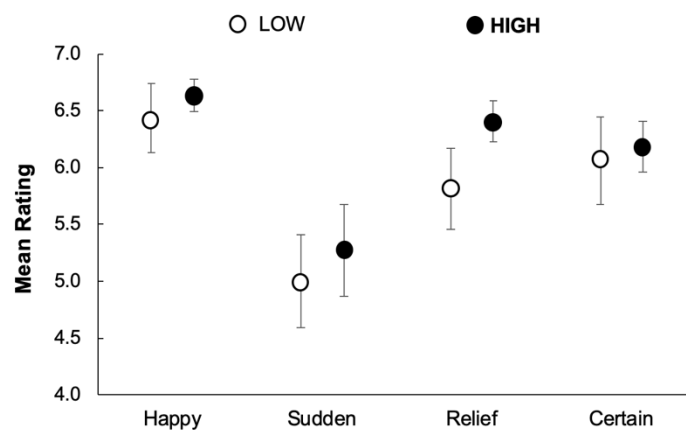
Figure 4. Mean rating along four experiential dimensions among successful participants in the low and high interactivity condition. Error bars are standard errors of the mean.

Individual Differences

We also profiled participants in terms of their visuo-spatial imagery skills with the Object-spatial Imagery Questionnaire (OSIQ) and the Image Control and Recognition Task (ICRT). The mean score on these scales for successful and unsuccessful participants in the low and high interactivity groups for both experiments are reported in Table 3. As participants were randomly allocated to either the low or high interactivity condition, we

expected no difference in OSIQ or ICRT scores between conditions, and indeed none were observed (all ps > .295). While successful participants generally scored higher than unsuccessful participants on both scales and in both interactivity conditions, these differences were not statistically significant save for the difference in ICRT scores in the low interactivity condition, $t(33) = 2.835$, $p = .008$, and in the high condition, $t(31) = 2.356$, $p = .025$, in Experiment 2.

Table 3

Mean OSIQ and ICRT Scores in the Low and High Interactivity Group (as a Function of Success and Overall) in Experiments 1 and 2 (along with the Standard Deviation of the Mean).

Experiment 1		OSIQ		ICRT	
Group		*M*	*SD*	*M*	*SD*
Low					
	Successful	91.15	11.65	6.40	2.95
	Unsuccessful	83.64	6.86	4.91	2.51
	Overall	88.48	10.73	5.87	2.85
High					
	Successful	90.10	13.19	5.33	2.18
	Unsuccessful	85.13	15.13	4.50	4.21
	Overall	88.72	13.66	5.10	2.82
Experiment 2					
Low					
	Successful	92.25	15.75	7.75	3.15
	Unsuccessful	90.47	13.78	4.68	3.21
	Overall	91.29	14.51	6.09	3.50
High					
	Successful	93.87	13.55	6.57	2.63
	Unsuccessful	92.90	17.87	3.90	3.73
	Overall	93.65	14.93	5.84	3.20

Qualitative Observations: Material Traces of Thinking

All participants were video recorded as they engaged with the triangle of coins problem. We then transcribed the high interactivity videos in terms of individual movements, their timing (from which we could also note the latency per move), and the resulting change in the problem configuration. The tokens could be dragged using the touchscreen interface to any cell on the chequerboard. For the purposes of the current study, we were interested in changes in the array, that is the external representation of the problem, rather than embodied actions or motor activity. Although we recognize that such moves would play a role in problem solving, the granularity of analysis required was not necessary to answer our research question. A move was therefore defined as a change in the array. The time of the move was taken when a participant moved a coin to a new position and took their finger off it. In the process, we recovered the material traces of thinking for each participant. These traces map the dynamic physical re-configuration of the problem as participants labour a solution. These material traces also reveal the temporal dimension of thinking, and the multi-scaled topography over which the journey to discovery proceeds. This scalar relief is illustrated by the wide variation in move latencies (clearly revealed in the case studies described below). The aggregate analyses reported above (and illustrated in Figure 3) suggest that the discovery of the relevance of the corner coins and their position was gradual, developing from extensive experimentation with different coins and different location on the board. In this section, we increase the granularity of the analysis by looking at the material traces of thinking from individual participants. These material traces unveil several telling features of problem-solving activity namely (i) the incorrect initial strategy or heuristic to solve the problem (and recurring incorrect configurations), (ii) the degree of perseverance and incomprehension of the task goal and constraints, and (iii) the genesis of the solution through the gradual configuration of the coins into the correct pattern.

A

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4									
5		Q		J		Z		W	
6			S		X		T		
7				L		Y			
8					R				
9									

B

	A	B	C	D	E	F	G	H	I
1									
2					R				
3				L		Y			
4			S		X		T		
5		Q		J		Z		W	
6									
7									
8									
9									

C

	A	B	C	D	E	F	G	H	I
1									
2					R				
3									
4			S				T		
5		Q		J		Z		W	
6					X				
7				L		Y			
8									
9									

D

	A	B	C	D	E	F	G	H	I
1									
2									
3					X				
4				L		Y			
5		Q		J		Z		W	
6			S				T		
7									
8					R				
9									
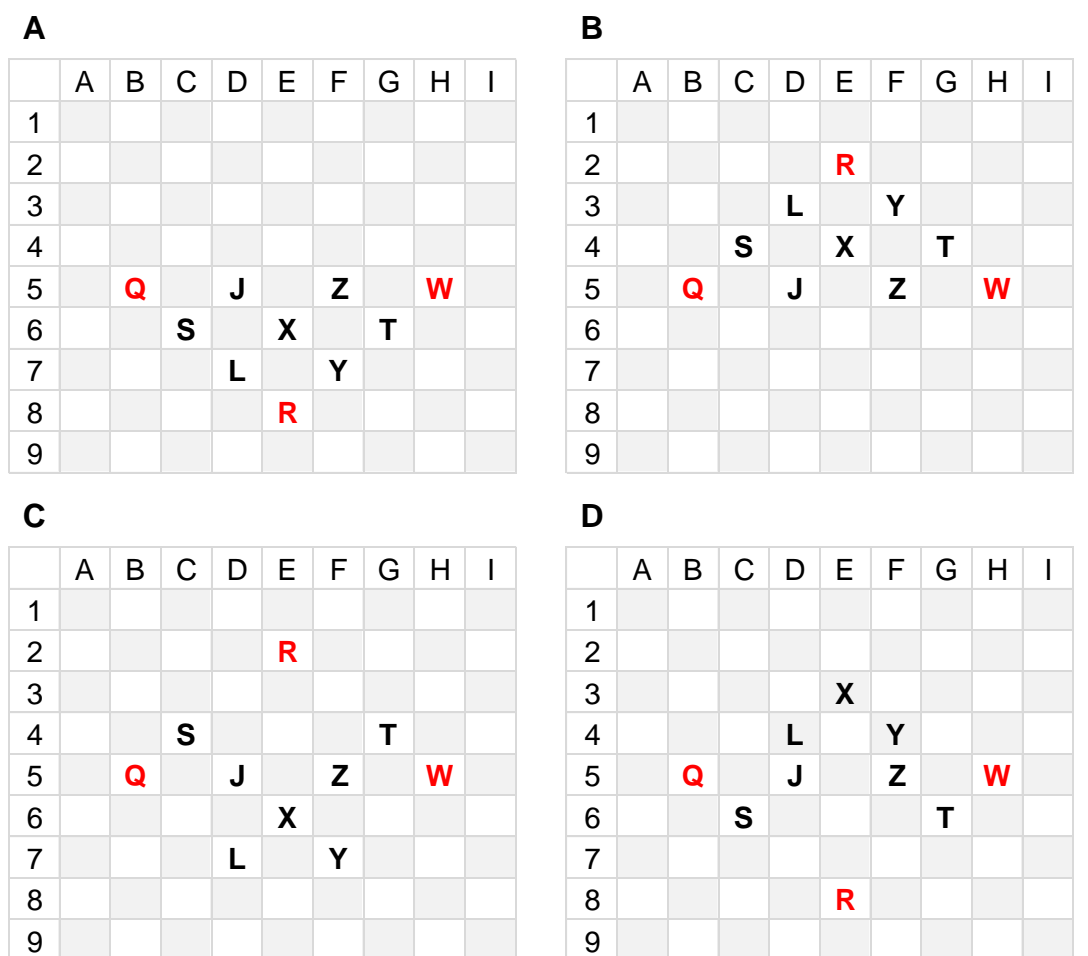
Figure 5. Panel A: Initial problem configuration (or start state). Moves to migrate coins 'north' cannot result in a solution: six coins moved up to produce a north-pointing triangle (Panel B, participant 35, Experiment 2, 12th move); Panels C (participant 54, Experiment 1, 44th move) and D (participant 65, Experiment 1, 5th move) illustrate non-productive configurations that respect the three-coin only constraint. Note the corner coins are colour coded here for ease of interpretation; they were not colour coded in this manner for the participants.

Initial Strategy: Why the Triangle of Coins Is a Difficult Problem

First, these traces reveal that the triangle of coins problem is a difficult so-called insight problem because most participants at some point during the task aim to migrate coins north to create a triangular shape that points up. That is, rather than moving coins south (from row 5 in Panel A of Figure 5) to form a new base along row 7, participants leave row 5 untouched and move coins above it to recreate a triangular shape pointing up. No solution can be evinced from this strategy because either too many coins must be moved up to create a

pseudo solution (panel B, Figure 5), or if only three coins are moved up, the resulting shape is insufficiently triangular to count as a solution (panels C and D, Figure 5). Thus, solving the triangle of coins problem involves abandoning this strategy[5].

Dead Ends and Incomprehension

These material traces also help us identify the degree to which participants persevere in exploring unproductive paths to solution as well as participants' understanding of the challenge to solve the problem. For example, participant 57 in Experiment 1, created the pattern illustrated in Figure 6 panel A 13 times, moving to a conceptual cul-de-sac time and again, resetting the board and moving along the same unproductive trajectory (a type of impasse as defined by Fleck & Weisberg, 2013). In turn, changes to the configuration of coins illustrated in panel B (Figure 6, participant 51, Experiment 2) reflect a deep incomprehension of the goal of the task and the constraints to achieve it[6]

A

	A	B	C	D	E	F	G	H	I
1									
2					R				
3				J		Z			
4									
5		Q						W	
6			S		X		T		
7				L		Y			
8									
9									

B

	A	B	C	D	E	F	G	H	I
1									
2	L								J
3									
4	Q								R
5									
6	S								T
7									
8	W								X
9				Y		Z			

Figure 6. Panel A: Recurring configuration created by participant 57, Experiment 1 (11th of 13 iterations); panel B, configuration created by participant 51, Experiment 2, after 40 moves.

---

[5] The task instructions might have encouraged participants to move coins up since the triangle is introduced as an arrow - that is, it is converted from a geometric shape into an iconic sign. The goal is to make the arrow point up. Thus, the signification of the arrow must change: the point must go from below to above. There is a semiotic logic in moving coins in the same direction as the new direction of the arrow: The coins migrate north because the arrow must point north. Different task instructions might reduce the likelihood of these initial moves. We thank Paul March for this point.

[6] A more systemic explanation could also be offered: if the application was programmed to respect the rules imposed by the problem and would not allow participants to continue moving the coins after they moved three, such deep incomprehension would likely not manifest.

Genesis of Solution: Gradual Development of the Correct Configuration

We developed an index termed the migration ratio to measure participants' progress in shifting the shape of the triangle to arrive at the correct configuration. The migration ratio is the ratio of coins on row 7 over all coins above on the grid: in the start state the ratio of coins in row 7 over all coins above is 2/7 or .29 (see Start panel in Figure 7). The goal state involves widening the base of the new triangle on row 7, thus migrating down the two corner vertices (coins Q and W) on that row, resulting in a migration ratio of 4/6 or .67 (see Goal panel in Figure 7). Calculating the migration ratio after each move mapped participants' efforts to create a configuration that would more productively evince a solution. In contrast to the corner move analysis reported earlier, the migration ratio more fully captures the distribution of the coins and their arrangement on the grid, while the quartile corner move analysis only focuses on the movement of three coins without capturing their location on the board as such.

START

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4									
5		Q		J		Z		W	
6			S		X		T		
7				L		Y			
8					R				
9									

GOAL

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4					R				
5				J		Z			
6			S		X		T		
7		Q		L		Y		W	
8									
9									

Figure 7. Migration ratio: Number of coins along row 7 over the number of coins on rows above. The migration ratio at the start state equals 2/7 or .29; it equals 4/6 or .67 at the goal state.

A low migration ratio index indicates that a participant attempts to migrate coins up, while a higher migration ratio is indicative of a strategy where coins are moved down to widen the base of the triangle. This is imperfect since, as pointed out by a reviewer, many clearly incorrect configurations can push up the ratio. Yet, however blunt a measure, it

captures a crucial change in strategy, namely that coins should be moved down rather than up. Thus, the ratio measures an important migratory change in movements, which signals that the problem is modified in a manner that is more likely to result in a solution.
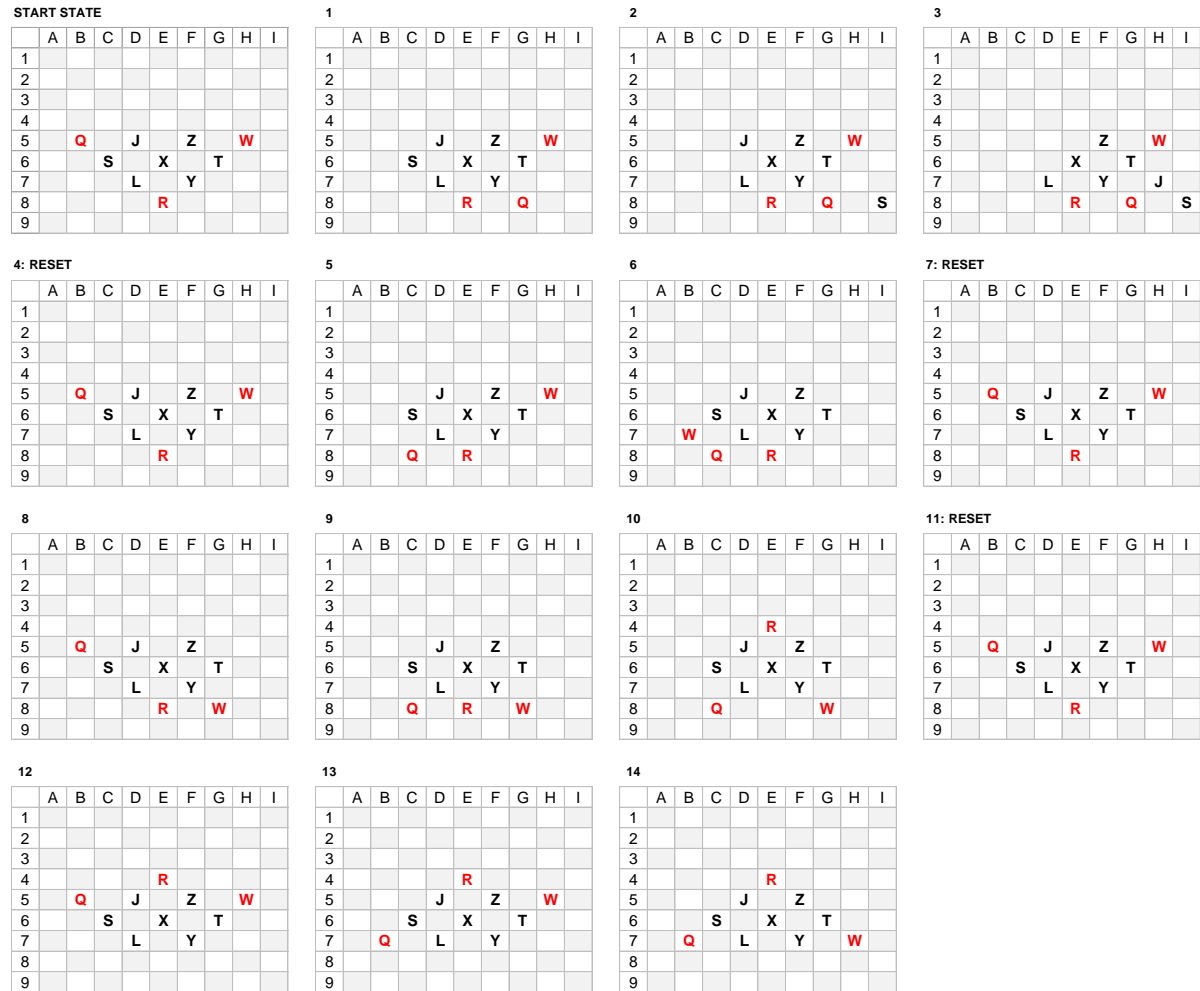
Figure 8. The configurations created by participant 14, Experiment 1, before announcing the correct solution.

Figure 8 illustrates the 14 moves to solution from participant 14 in Experiment 1, while Figure 9 (panel A) captures the participant's trajectory to solution as a function of changes to the migration ratio (along the primary y axis) and latency per move (along the secondary y axis). The migration ratio oscillates considerably as do the move latencies for the first 8 moves. The solution latencies vary much less as the board is reset (move 11): They stabilize just before dropping sharply with the final move which coincides with coin W moved to the base of the new triangle in the 14th frame. The final move latency suggests that at that point

in time, the participant could clearly anticipate the answer; however, we would argue that she

did not before that point. If the participant had a clear mental representation of the answer by

the time she reset the board in the 11th frame, we would have seen a steep decline in move

latency (in fact the participant might have simply opted to submit an answer at this point, but

she did not; note that no successful participants in the high interactivity condition of both

experiments submitted an answer before physically configuring it[7]). However, the move

latency went up slightly in the 12th frame, to then decline slightly in the 13th frame before the

final move that took one second. The reciprocal mapping between the participant's mental

anticipation of the solution and the physical model of the solution aligned gradually. The

physical reification of the solution and the participant's mental representation of it are aligned

with the last move but not before: in other words, the solution is enacted, distilled through

action, rather than implemented.

The gradual alignment of a participant's appreciation of the solution with a physical

model of the solution is illustrated in Figure 10 (participant 33, Experiment 2; the dynamic

changes in the migration ratio and latency per move for this participant is illustrated in Figure

9, panel B). These 16 frames correspond to the 16 moves that lead to creation of the right

solution (the participant will go on to reset the board and recreate the solution a second time,

a phenomenon observed with seven participants in Experiment 2, but not once in Experiment

1, presumably as a response to the one-announcement constraint). More productive

configurations are created from the reset following the 45th move: corner coins from row 5

are migrated down, initially along row 8. Coin R is moved up in the 49th frame. There ensued

a series of small changes to this configuration, along with two resets (frames 52 and 57), that

eventuated in the solution. Thus, the right configuration was gradually tweaked: once the

---

[7] With an interesting exception: participant 39 in Experiment 2 simulated moving the corner coins Q and W
down to project a new base simultaneously with two hands, then submitted the answer without actually
physically constructing the solution on the interface.
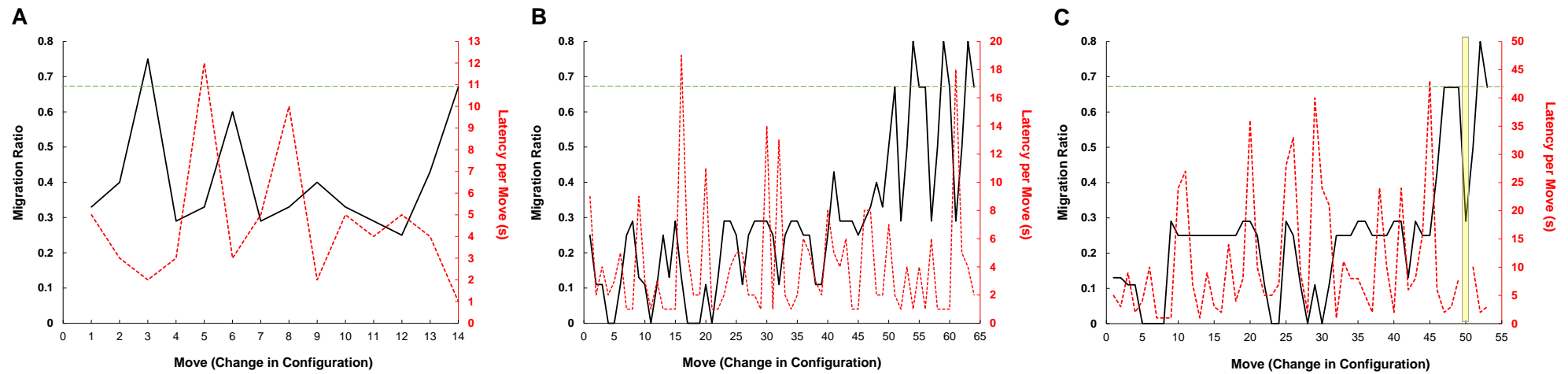
Figure 9. Changes in the migration ratio (along the primary y axis) and latency per move (along the secondary y axis) for participant 14, Experiment 1 (panel A); participant 33, Experiment 2 (panel B), and participant 7, Experiment 1 (panel C). The horizontal line along the upper portion of the graph indicates the target migration ratio that corresponds to the correct configuration of the coins; the shaded vertical section in panel C marks the participant's incorrect announcement.
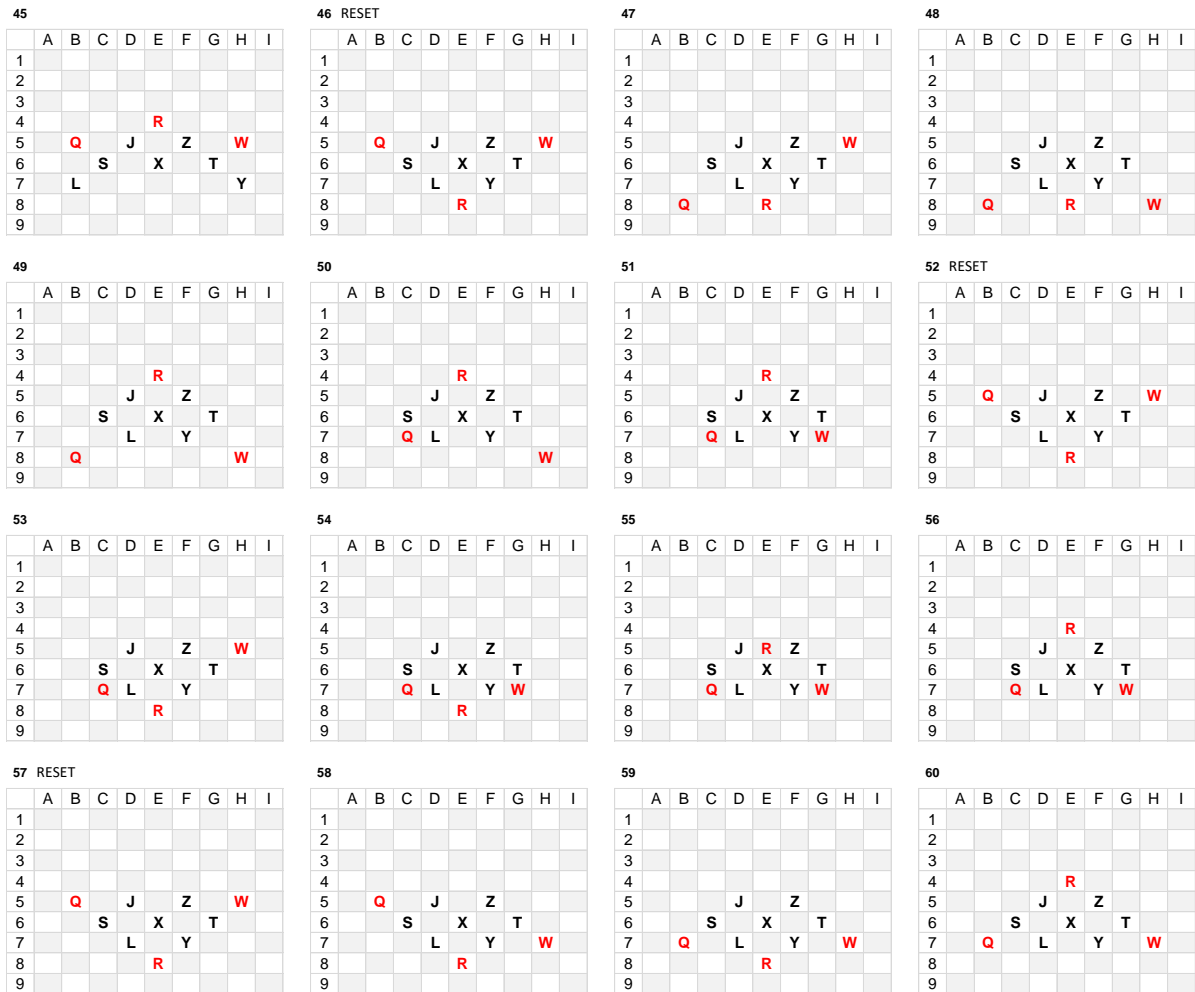
Figure 10. Participant 33 (Experiment 2): gradual change to the problem configuration resulting in the problem solution in 16 frames.

corner coins from row 5 were moved down to widen the base of a new triangle, the

participant invested in this more promising configuration (after toiling for 40 frames moving

coins up rather than down) and carefully explored different positions of the vertices to create

a triangle pointing up: The solution was enacted slowly rather than reflecting a sudden

insight. The participant did not announce the solution at the 60th frame; rather she reset the

board and re-created the solution a second time, and then announced the correct answer.

Finally, the progression of the migration ratio to the correct configuration for

participant 7 (Experiment 1) is illustrated in panel C in Figure 9. The participant labours

through unproductive configuration for most of the session, as reflected in migration ratio

values that vary between 0 and .3. The latency per move data for this participant show that

changes to the problem configuration were particularly slow between resets on move 25 and move 35, a work period marked by a particularly unproductive series of changes to the problem configuration (see top half of Figure 11, illustrating moves 26-28). Move latencies drop considerably after move 35 and a new peak in latency coincides with move 45: move 46 involves migrating the right corner vertex down to row 7, followed by moving the left corner vertex down (move 47; see bottom half of Figure 11). This is a much more informative configuration; indeed, P7 solved the problem with her next few moves, although not before making an incorrect announcement based on the configuration illustrated in the 48th frame (an incorrect solution involving the movement of 4 coins).

26

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4					R				
5		Q		J		Z		W	
6			S		X		T		
7				L		Y			
8									
9									

27

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4				L		R			
5		Q		J		Z		W	
6			S		X		T		
7						Y			
8									
9									

28

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4				L		R		Y	
5		Q		J		Z		W	
6			S		X		T		
7									
8									
9									

46

	A	B	C	D	E	F	G	H	I
1									
2				X					
3									
4									
5		Q		J		Z			
6			S		R		T		
7				L		Y		W	
8									
9									

47

	A	B	C	D	E	F	G	H	I
1									
2				X					
3									
4									
5				J		Z			
6			S		R		T		
7		Q		L		Y		W	
8									
9									

48

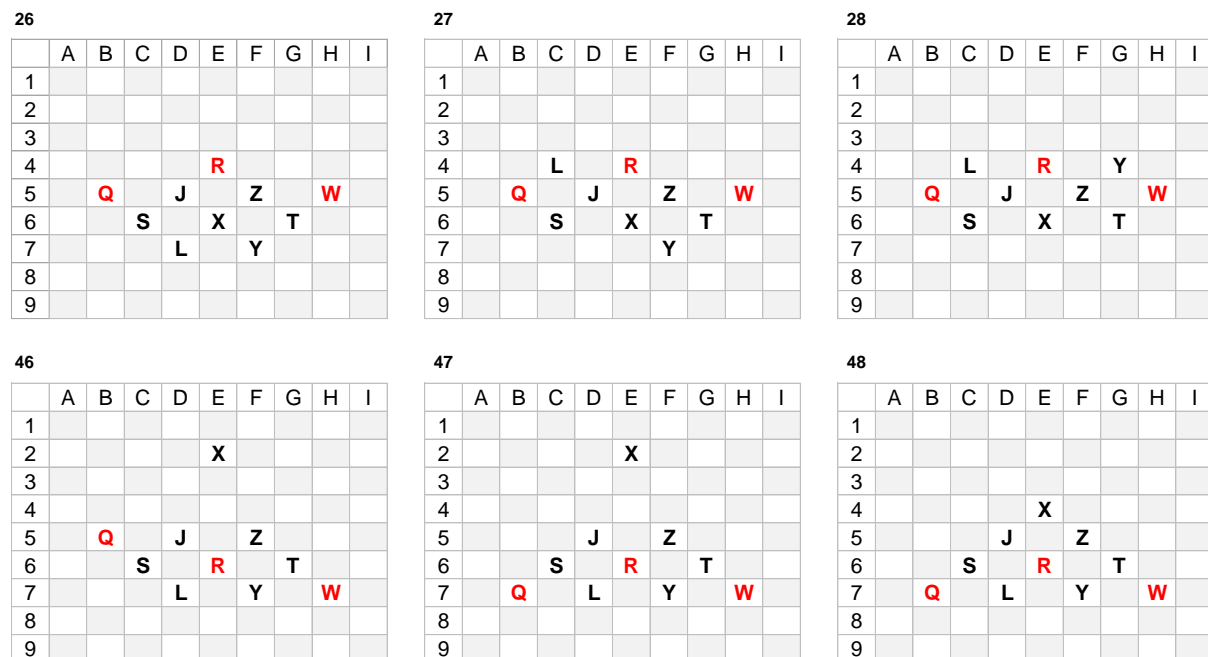|	A	B	C	D	E	F	G	H	I
1									
2									
3									
4					X				
5				J		Z			
6			S		R		T		
7		Q		L		Y		W	
8									
9									

Figure 11. Participant 7 (Experiment 1): Unproductive configuration in moves 26-28 coinciding with relatively slow latency per move (top half); moves 46-48 (bottom half) illustrate a more productive configuration enacted with relatively faster moves.

Metcalfe's Rotational Insight

Metcalfe (1986) originally suggested that the triangle of coins could be solved in a rotational manner eliciting a 'rotational insight' (see Figure 12). However, Chronicle, MacGregor and Ormerod (2004) suggested that the problem can also be solved by moving the top coins down to form a new base ('transformation across the median', p.23).
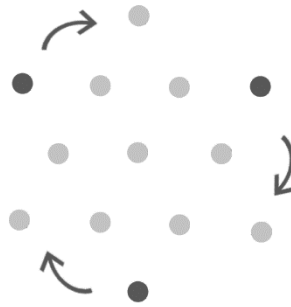
Figure 12. The rotational solution as suggested by Metcalfe (1986).

Chronicle et al. (2004) used physical coins to structure the problem-solving attempts and also filmed and coded the participants' solution attempts. When they charted the solution attempts, they suggested that the transformation across a median was the more common move (8 of 34 exhibited a rotational pattern whereas 19 used this median transformation pattern; Chronicle et al., p. 19). We observed similar patterns here: In 17 out of 20 correct solution in Experiment 1 and 16 out of 21 in Experiment 2 involved using this transformation. Chronicle et al. suggested that the median transformation solution represents the result of a hill climbing heuristic and should lead to lower levels of insight. This hypothesis is not supported by our data: the overall subjective rating of insight for both strategies was broadly similar. The average reported insight (on the basis of the sum of the ratings on the four dimensions) for the rotational strategy was 24.17 (SD = 2.97) and for the median transformation was 24.29 (SD = 12.47), $t(21)$ = -.008, p = .934. Indeed, the move by move coding of the data would suggest that characterising one of these two types of solution as more reflective of a sudden restructuring would not reflect our data. Take participant 15 (Experiment 2): the final moves demonstrate how the rotation solution unfolds itself through an externalisation of different moves rather than as a sudden internal restructuring as theorised by both Metcalfe and Chronicle (see Figure 13). This suggests that solution types

which have been previously characterised as demonstrating purer forms of insight (e.g. rotational vs median transformation) are also distilled by action.
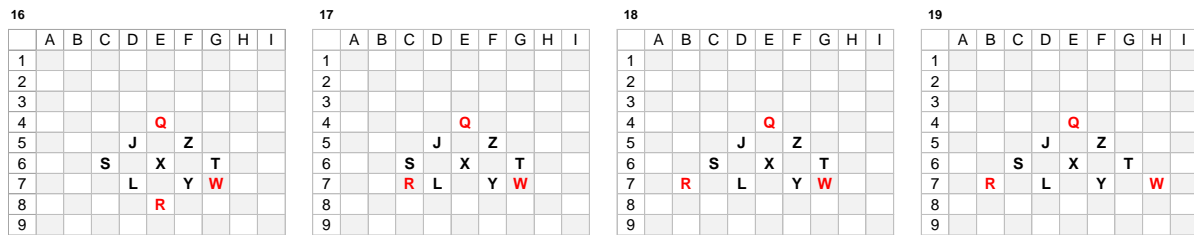
16

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4					Q				
5				J		Z			
6			S		X		T		
7				L		Y	W		
8					R				
9									

17

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4					Q				
5				J		Z			
6			S		X		T		
7			R	L		Y	W		
8									
9									

18

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4					Q				
5				J		Z			
6			S		X		T		
7		R		L		Y	W		
8									
9									

19

	A	B	C	D	E	F	G	H	I
1									
2									
3									
4					Q				
5				J		Z			
6			S		X		T		
7		R		L		Y		W	
8									
9									

Figure 13. The final four moves for Participant 15, Experiment 2.

This incremental movement to solution reflects similar findings in other studies which have moved beyond simple aggregated solution rates. Bilalić, Graf, Vaci and Danek (2019) tracked eye movements across solutions of matchstick arithmetic problems and their data suggest a monotonic, gradual temporal dynamics for the majority of participants (although it is important to note that sudden discontinuities in looking patterns were also found for a quarter of the participants). Our finely grained analysis lends support to the data reported in Bilalić et al. as well as confirms the aggregate analyses reported above: the appreciation of the importance of the corner coins arises gradually over time.

Discussion

The triangle of coins is a hard problem in the absence of the cognitive scaffolding offered by an interactive task environment. This is especially true if participants are prevented from offering multiple incorrect announcements as they are working on the problem. Participants' solution rate in a high interactivity task environment was constant at the 70% mark in both experiments, although the one-announcement constraint in Experiment 2 encouraged some participants in that condition to physically create the solution more than once (and contributed to the longer solution latencies in that experiment). Fleck and Weisberg (2013) also gave participants artefacts to construct physical models of the solution with the triangle of coins problem and reported a similar solution rate over a 10-min period (viz. 69%; see also Vallée-Tourangeau, 2017). Interactivity is neither necessary to solve this

problem nor a panacea, however: nearly half of the participants in the low interactivity condition managed to solve the problem without multiple announcements and nearly a third failed to announce a correct solution in the high interactivity condition. Furthermore, the recurring creation of unproductive patterns (as illustrated in Figure 6, panel A) reflected a deep conceptual impasse that might have been exacerbated through the habitual movements that physically reinstated it. The post-task comprehension question in Experiment 2 revealed that 80% of the unsuccessful participants in the high interactivity condition could not recognize the correct solution—compared to 56% among unsuccessful low interactivity participants—which might reflect some interference caused by the engrained non-solution configurations that were so often recreated.

The dominant hypothesis in research into interactive problem solving is that interactivity confers a consistent benefit. It is tempting to think that extending the cognitive system must necessarily strengthen it. This is often the case where the system acts as a simple offload but it is not a necessary quality of the extended mind (Wilson, 2014) and indeed as the research programme into materially and socially extended cognition matures, it is becoming clear that there is not a straightforward additive relationship which can be easily parsed through experimental study (see Vallée-Tourangeau & Vallée-Tourangeau, 2020b). For example, work on mental arithmetic suggests that interactivity offers a benefit for complex additions but not for simpler ones (Vallée-Tourangeau, 2013) suggesting that the cognitive costs of interactivity could outweigh their benefits. The quality of coupling between mind and world is important: Gilbert (2015) found that not all participants made use of offloading to support memory although it was beneficial to do so.

Moreover, beyond the difficulty of mapping idiosyncratic open systems that evolve in more fluid and complex task ecologies, insight tasks cannot be solved through simple step by step moves. In this instance, whereas interactivity broadly confers a boost to tasks requiring

working memory by reifying thoughts and expanding the mental workspace, a task with multiple unhelpful pathways may actually be made more difficult through interactivity: The reification of unhelpful solution attempts may make it harder to disregard these and lead to a reduction in performance (see Ross and Vallée-Tourangeau, 2020 where participants in a word production task created words by moving physical letter tiles, sometimes found it harder to generate new words when a word was in front of them rather than mentally conjured; the mental fixation presumably less tangible and then also easier to disregard).

Impasse, Heuristics and Restructuring

Fleck and Weisberg (2013) employed verbal protocols to identify several components of problem-solving including rates impasse, the types of heuristics employed and the restructuring of a mental representation. The latter may be endogenously generated through the reasoner's top-down analysis of the problem (termed "conceptually driven restructuring") or exogenously suggested through unintentional changes to the external configuration of the problem (see p. 453; termed "data driven restructuring"). The obvious benefit of verbal protocols is the potential window they offer on the thinking that accompanies problem-solving efforts. In the experiments reported here, we simply do not know what heuristics, moments of impasse or restructuring might have characterized participants' mental effort in the low interactivity condition. Instrumentalizing a problem-solving task as we have done here provides data that enable the physical mapping of the trajectory to solution (or impasse). The coding in the high interactivity condition of the changes in the external configuration of the problem along with their timing provided concrete illustrations of the dynamic changes in the physical presentation of the problem. The qualitative coding revealed instances of impasse, as reflected by unproductive configurations (see Figure 6) and long latencies (see panel C, Figure 9), although no complete behavioural cessation. One of the interesting features of high interactivity is that the task environment affords movement, and it is through

movement that the environment provides information to guide and constrain problem solving activity.

There is evidence of heuristics and heuristic switching in these qualitative data: Moving coins north as a rough means-end analysis heuristic, and very likely lots of trial and error actions and random exploration ("the simplest heuristic", Fleck & Weisberg, 2013, p. 452). As for evidence of restructuring, well, all we can say is that eventually the physical configuration of the problem was restructured into the correct configuration; whether this was preceded by data-driven or conceptually-driven restructuring, or indeed whether restructuring—as realised through one of Ohlsson's (1992) three mechanisms, namely elaboration, reencoding and constraint relaxation—occurred at all, remains to be determined. Adding verbal protocols to a procedure such as ours might more clearly identify the movements that are clearly motivated by a hypothesis, rather than reflecting exploration that is not guided by a specific plan.

Both quantitative aggregate data and the qualitative coding revealed that the correct configuration was often gradually distilled through changes in the appearance of the triangle. We would argue that a gradual solution is anything that takes more than three physical moves. This may appear a facile or even facetious point, but we do not intend it in this manner. The solution to the problem is the historical aggregate of all the moves that preceded its implementation. Seen from this perspective, most of our participants discovered the solution gradually. The separate phases of the problem solution are reliant on the ones that came before – there is a complex enmeshing of enacting the impasse and trialling unhelpful strategies which leads to recognizing the correct solution. We suggest that viewing each phase separately and isolating it from what comes before would cloud our understanding of the process of discovery. Having said this, there are plenty of discontinuities—which we also

take to support the multi-scalar argument—and these are abundantly illustrated in the latency peaks and troughs of Figure 9.

We propose that the process of discovery is more productively conceived in transactional terms, where the mental representation of the solution and the physical reification of the solution mutually shape their development. As mentioned above, Fleck and Weisberg (2013) distinguish between two types of restructuring, conceptually driven and data driven. Yet, both types of restructuring as described by these authors can occur through the manipulation of a physical presentation of the problem, and all the cognitively felicitous consequences that this entails. It is not clear to us how useful this difference is, or indeed how precisely the difference between these two types of restructuring can be established. From a transactional perspective, the intention of an action is difficult to attribute to either an agent or the environment; executive control is an emergent and distributed consequence of the agent-environment coupling. That is, the task environment offers a shifting set of action affordances which, once enacted, results in in dynamic changes to the physical configuration of the problem and so on. Thus, when participants abandon migrating coins north and start exploring consequences of migrating the base down, it is not clear whether this reflects a restructuring of a data-driven or conceptually driven kind (or any mental restructuring at all). The very gradual adjustment illustrated in Figure 10 suggests that a participant's appreciation of the solution was aligned with and motivated by changes in the physical configuration concurrently. What was unequivocally restructured was the physical configuration of the problem elements (viz. the "coins") that gradually approximated the solution, and which offers a reflection of the participant's own understanding of the solution.

Outsight. The coding of participant 33 (as illustrated in Figure 9, panel B, and Figure 10) revealed that however close to the correct answer a created configuration might be, a participant might not *see* the right answer. This suggests that some of the participants might

have constructed the correct configuration and only recognized the construction as offering a model of the solution after having done so. Thus, 'insight' in this case is perhaps better understood as 'outsight' (Vallée-Tourangeau & March, 2019). Participant 61 in Experiment 2 provides a clear example of this phenomenon (the video can be accessed here: https://osf.io/hbs3f/?view_only=394d3ba358974462817dd38ba4280cdf). The moment of outsight can be observed at 0:00:07: After moving the W coin down to row 7 to form the base of the triangle pointing up, the participant displays an exaggerated 'double take', with hands splayed open. Such evidence reveals the tight co-evolution of knowledge of the solution and the physical shape of it: the movements of participant 61 are not guided by a pre-formed plan of the solution, but rather knowledge of the solution is physically enacted and reified.

It is worth noting an interesting phenomenon observed with the ICRT, one also reported by Irving (2014) in her doctoral dissertation on the development and psychometric properties of the ICRT. Across both experiments, we noted over 100 instances of participants being able to name the figure described in the guided synthesis instructions after drawing it, but not before. From our perspective, the phenomenon is interesting, and might deserve more systematic attention, because it illustrates a form of outsight: People's ability to know what they are thinking is only fully realized when they create an external representation of the content of their thoughts. Participants' behaviour working on the triangle of coins in the high interactivity condition might also reflect a similar phenomenon and underscores the importance of working with dynamic and modifiable external representations in problem solving.

Contribution to the Business-as-usual Versus Special Processes Debate

Finally, let us reflect on how interactivity and the findings from the present experiments inform the current debate in the insight problem solving literature between the so-called business-as-usual and special processes perspectives. These perspectives, whatever

predictions they make about the relative merit of unconscious inferences or conscious

deliberate analyses, are couched in mentalist terms, in turn predicated on a deeply entrenched

form of dualism. By this we mean that for theorists of both camps, new ideas are wrought

through the mental manipulation of abstract representations, and hence the ontological locus

of creativity and discovery is in the person, or more specifically, the person's head. In

contrast to this form of methodological individualism, our approach is aligned with

methodological interactivism (Bickhard, 2009). An interactivist perspective casts the origin

of new ideas as an emergent property of relational processes; Steffensen (2017) refers to

interactivity as an 'ontological substrate', that is new ideas emerge out of the interaction with

the world. In our experiments, interactivity fosters the iterative attunement of a participant's

proto solution with the correct physical configuration of the problem. One might argue that

the interface employed in the experiments reported here artificially encouraged this form of

trial and error, and in that respect perhaps such an interface clouds the true underlying

cognitive mechanisms which this task is designed to elicit, the mechanisms which will scale

up to support our understanding of more complex problem outside the laboratory

environment. However, we contend that such a high interactivity task environment offers a

much more likely candidate for the process of scaling up mechanisms of creative cognition

than a task environment in which participants are restrained from interacting with a physical

environment. For example, science studies (e.g., Latour's [1987] *Science in action*) brim with

examples of the messy iterative alignment of all manner of 'non-human actants' (to borrow a

term from Latour) with provisional solutions in the reification of meaning, new discoveries

and ideas. Our findings do support the kind of progressive gradualism so often illustrated and

argued by Weisberg both in his case studies of scientific and artistic innovation (e.g.,

Weisberg, 2006) as well as in his laboratory work (e.g., Fleck & Weisberg, 2013): Local

changes and cumulative minor discontinuities with the past can culminate into larger discontinuities that can be deemed innovative and creative.

Conclusion

Köhler's observations of Koko's efforts to solve the box test might not have been so exceptionally 'curious': There is much evidence in the data reported here of the gradual discovery of the configuration that offered a solution to the triangle of coins, a solution that slowly dawned on the majority of our participants (to adapt Köhler). The instrumentalization of a classic insight problem—by presenting the coins on a labelled grid and filming participants—enabled the granular coding of the changes in the physical configuration of the problem from which we could create material traces of the participants' exploration, mapping out their journey to discovery. A task environment that encourages first order problem solving and interactivity lends itself to that kind of qualitative analysis and in the process reveals the rugged scalar relief of thinking.

References

Austin, J. H. (1979). The varieties of chance in scientific research. *Medical Hypotheses, 5*, 737-742. doi: 10.1016/0306-9877(79)90035-5

Bickhard, M. H. (2009). Interactivism: A manifesto. *New Ideas in Psychology, 27*, 85-95. doi: 10.1016/j.newideapsych.2008.05.001

Bilalić, M., Graf, M., Vaci, N., & Danek, A. H. (2019). The temporal dynamics of insight problem solving: Restructuring might not always be sudden. *Thinking & Reasoning*. doi: 10.1080/13546783.2019.1705912

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology, 20*, 239–263. doi: 10.1002/acp.1182

Chronicle, E. P., MacGregor, J. N., & Ormerod, T. C. (2004). What makes an insight problem? The role of heuristics, goal conception, and solution recoding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 14-27. doi: 10.1037/0278-7393.30.1.14

Chuderski, A., Jastrzębski, J., & Kucwaj, H. (2020). How physical interaction with insight problems affects solution rates, hint use, and cognitive load. *British Journal of Psychology*. doi:10.1111/bjop.12442

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2014). It's a kind of magic—what self-reports can reveal about the phenomenology of insight problem solving. *Frontiers in Psychology*, *5*, 1408. doi: 10.3389/fpsyg.2014.01408

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

Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research and applications*. Cambridge, MA: MIT Press.

Fiore, S. M., & Wiltshire, T. J. (2016). Technology as teammate: Examining the role of external cognition in support of team cognitive processes. *Frontiers in Psychology, 7,*1531. doi: 10.3389/fpsyg.2016.01531

Gilbert, S. J. (2015). Strategic offloading of delayed intentions into the external environment *Quarterly Journal of Experimental Psychology, 68,* 971-992. doi:10.1080/17470218.2014.972963

Glăveanu, V. P., Lubart, T., Bonnardel, N., Botella, M., Biaisi, P.-M. de, Desainte-Catherine, M., … Zenasni, F. (2013). Creativity as action: Findings from five creative domains. *Frontiers in Psychology, 4*, 176. doi: 10.3389/fpsyg.2013.00176

Irving, L. (2014). *Creativity, imagery and schizotypy: An exploration of similarities in cognitive processing*. Unpublished doctoral dissertation, Middlesex University, UK.

Irving, L., Barry, R., LeBoutillier, N., Westley, D. (2011). The image control and recognition task: A performance-based measure of imagery control. *Journal of Mental Imagery, 35*, 67-80.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Malafouris, L. (2020). Thinking as 'thinging': Psychology with things. *Current Directions in Psychological Science*, *29*, 3-8. doi: 10.1177/0963721419873349

Metcalfe, J. (1986). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 623-634. doi: 10.1037/0278-7393.12.4.623

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* (Vol. 1, pp. 1-44). Harvester Wheatsheaf.

Ross, W. (2020) Serendipity. In V. Glăveanu (Ed) *The Palgrave encyclopedia of the possible*. doi:10.1007/978-3-319-98390-5_47-1

Ross, W., & Vallée-Tourangeau, F. (2020). Catch that word: Interactivity, serendipity and verbal fluency in a word production task. *Psychological Research*. doi:10.1007/s00426-019-01279-y

Steffensen, S. V. (2017). 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* (2nd edition, pp. 85–113). Dordrecht: Springer.

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

Vallee-Tourangeau, F. (2014). Insight, interactivity and materiality. *Pragmatics & Cognition, 22*, 27–44. doi: 0.1075/pc.22.1.02val

Vallée-Tourangeau, F. (2017). Interactivity and ego depletion in insight problem solving. In G. Gunzelmann, A. Howes, T. Tenbrink, & E.J. Davelaar (Eds.), *Proceedings of the Thirty Ninth Annual Conference of the Cognitive Science Society* (pp. 1248-1253). Austin, TX: Cognitive Science Society.

Vallée-Tourangeau, F., & March, P. L. (2019). Insight out: Making creativity visible. *Journal of Cognitive Behavior*. doi: 10.1002/jocb.409

Vallée-Tourangeau, F., & Vallée-Tourangeau, G. (2020a). Insight. In V. P. Glăveanu (Ed.), *The Palgrave encyclopedia of the possible*. Springer.

Vallée-Tourangeau, F & Vallée-Tourangeau, G. (2020b). Mapping systemic resources in

> problem solving. *New Ideas in Psychology*. doi: 10.1016/j.newideapsych.2020.100812

Weisberg, R. W. (2006). *Creativity: Understanding innovation in problem solving, science,*

> *invention, and the arts*. Wiley.

Wilson, R. A. (2014). Ten questions concerning extended cognition. *Philosophical*

> *Psychology, 27,* 19–33. doi: 10.1080/09515089.2013.828568