Title
Representational Effects in a Rule Discovery Task

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN
1069-7977

Authors
Peshkin, Leonid
Savova, Virginia

Publication Date
2005

Peer reviewed

Representational Effects in a Rule Discovery Task

Frédéric Vallée-Tourangeau, Andrea Krüsi Penney and Teresa Payton
Psychology Research Unit, Kingston University
Kingston-upon-Thames, UNITED KINGDOM KT1 2EE
f.vallee-tourangeau /a.kruesi /t.payton@kingston.ac.uk

Abstract

In the Wason (1960) rule discovery task reasoners must infer a rule that governs the production of number sequences. The task instructions are designed such that reasoners' initial hypotheses are invariantly narrower than the correct rule. The inferential challenge lies in discovering the scope of these initial hypotheses. The traditional task departs from real-world hypothesis testing in at least one significant respect: The task is never presented in a manner that offers a rich external representation of the problem. The study reported here examined representational effects by developing task isomorphs which offered either an external, physically manipulable, representation of the problem space or a graphical presentation of the simple linear relationships between adjacent numbers in a sequence. Compared to a control condition, these task isomorphs lead to a significantly higher incidence of successful rule discovery, and encouraged the creation of a significantly more heterogeneous set of number sequences. Further, control participants had to produce 50% more number sequences than participants who worked with a rich external representation before discovering the rule.

Keywords: Reasoning; distributed cognition.

Introduction

Wason (1960) created a simple inductive inference context which cast light on reasoners' ability to discard plausible but overly narrow hypotheses. Wason's 2-4-6 task consists of discovering the rule that governs the production of sequences of three numbers, or triples. To do so, participants create new number sequences that are then classified as either conforming or not to the rule. In Wason's original task the rule is 'any increasing sequence'. Before formulating their first test triple, participants are given the following crucial but misleading piece of information: the triple 2-4-6 conforms to the rule. This information implicitly identifies a set of triples that is narrower and more structured than the set of all conforming triples (Klayman & Ha, 1987). The salient features of the 2-4-6 example, namely evens and constant increments, constrain the initial hypotheses formulated by reasoners just as Wason (1960) intended. New sequences motivated by such initial hypotheses (e.g., 8-10-12) will unfailingly receive positive feedback since they are predicated on features that are sufficient but not necessary to produce triples that receive positive feedback. The challenge in this task thus lies in discovering the boundaries of such initially plausible hypotheses by formulating sequences that fall outside their scope, such as 1-19-33 for example.

Solving the 2-4-6 task is hard. In Wason's (1960) original study, 80% of participants failed to announce the correct rule at their first attempt. Subsequent replications have reported similarly low rates of success (e.g., Mahoney & DeMonbreun, 1977; Tweney, Doherty, Worner, Pliske, Mynatt, Gross, & Arkkelin, 1980; Wason, 1968). Two features characterise the hypothesis-testing output of most participants: indolence and narrow-mindedness (Vallée-Tourangeau, Austin, & Rankin, 1995). That is, participants produce few triples before announcing their best guess and those triples form a very homogenous set, which includes relatively few triples that increase by variable increments or that receive negative feedback. This rather grim reasoning profile has been consecrated into textbook wisdom (e.g., Poletiek, 2001; Schustack, 1988; Sutherland, 1992).

Two research avenues can be explored in an attempt to shed light on the poor rule-discovery performance in the 2-4-6 task. The first focuses on properties internal to the reasoners (cf. Stanovich & West, 2000). Wason himself enjoined future researchers to identify reasoners with a 'disposition to refute' (1960, p. 139). Vartanian, Martindale, and Kwiatkowski (2003) suggested that individual differences in creativity predicted successful rule induction in the 2-4-6 task. The second focuses on properties of the external environment. The aim is to determine whether variations in the physical presentation of the task encourage different degrees of diligence and creativity, and hence result in different degrees of success (Duncan, 1998). The research reported here stems from this latter perspective.

Representational Effects

Despite the simplicity of the 2-4-6 task, the inferential challenge of identifying the scope of a hypothesis and establishing its generalizability mirror features of real-world hypothesis-testing (Gorman, 1995). However, unlike real-world hypothesis testing the external representation of the dimensions of the triple space and the actual number sequences is nonexistent or impoverished. Rule discovery behaviour proceeds primarily on the basis of the reasoner's internal representation of the problem. The extent to which the representation of the problem is distributed between the reasoner's mind and the environment is limited to the written record of triples tested and feedback received. In this important respect the inferential context of the original Wason task is relatively atypical of real-world hypothesis testing. Scientific hypothesis testing, for example, is a process shaped by artefacts and methodologies favoured by researchers that encourage and constrain the nature of the hypotheses formulated and tested. Results of the scientific tests are also analysed and illustrated through a choice of representational media (e.g., a particular graphic format or diagram) that facilitates interpretation and best stimulates the formulation of new hypotheses (cf. Cheng, 1996). Thus, real-world hypothesis testing often proceeds from a much richer representation of the problem. The representation through which the problem is investigated and analysed is distributed over the reasoner's mental representation of the problem and its external representation structured by artefacts, cognitive or otherwise.

The work of Norman (1993) and Zhang (1997; Zhang & Norman, 1994) among others has clearly demonstrated that the manner with which a problem is externally represented shapes the cognitive strategies reasoners employ to solve it and hence their success at doing so. For example, Zhang (1997) looked at problem-solving behavior in the 'game of fifteen', a problem isomorphic to tic-tac-toe, where two players take turn selecting integers from 0 to 9 with the goal of being first to select three that sum to 15 (Simon, 1996, Ch. 5). While traditional tic-tac-toe offers a rich external representation of the problem, the game of fifteen proceeds primarily from an internal representation. As a result it fosters a much poorer and slower appreciation of the strategic imperatives for a draw than traditional tic-tac-toe. Such representational effects illustrate that problems isomorphic at some level of description may vary significantly in terms of the richness of the information embedded in their physical presentation. This variance can lead to significant differences in comprehension and problem-solving strategies.

Recently, Vallée-Tourangeau and Krüsi Penney (in press) investigated representational effects in the 2-4-6 task by offering a physically manipulable representation of the dimensions of the triple space. In a first experiment these dimensions were presented as three traditional 6-sided dice. The triple space in this task isomorph was thus defined in terms of 6^3 (or 216) triples composed of number sequences made up exclusively of positive integers ranging from 1 to 6. The substantial reduction in the size of the triple space might in itself have facilitated rule discovery. However, this did not appear to be the case. In the *1-6* control condition of Vallée-Tourangeau and Krüsi Penney's first experiment, 21% (4 out of 19) of participants announced the correct rule under such circumstances, compared to 20% in Wason (1960). Further, the hypothesis-testing profile of these participants gauged in terms of number of triples tested and their homogeneity was indistinguishable from the profile of the participants in Wason (1960). In contrast, participants in the *Dice* isomorph tested more triples of a more varied kind, including more nonincreasing triples, than in the *1-6* condition, and 66% (27 out of 41) announced the correct rule.

An external representation of the dimensions of the triple space made the number permutations perceptually salient seeding the triple generation process. Additionally, participants could physically manipulate these dimensions by rotating the dice and could produce new triples by moving the position of the dice. Alternatively, it could be argued that the familiarity of the dice as random number generating devices and their association with games of chance might have primed reasoners to think of the 2-4-6 task in fundamentally different terms (cf. Vanderhenst, Rossi, & Schroyens, 2002). However in Vallée-Tourangeau and Krüsi Penney's (in press) second experiment, reasoners generated triples using three sets of six hexagonal chips, each chip showing a circular array of open circles. A number from 1 to 6 was represented by having a corresponding number of shaded circles. None of the resulting patterns of dots matched the familiar patterns on any of the six sides of a traditional die. As with the *Dice* isomorph, participants working on the 2-4-6 task with the hexagonal chips tested a more heterogeneous set of triples and were significantly more likely to discover the rule than participants engaged in the traditional version of the task.

The research reported here extended the investigation of representational effects in the 2-4-6 task. In the current study, a larger triple space was employed by increasing the number range to 1-8 defining a triple space composed of 512 possible triples. In a first experimental condition, 8-sided role playing dice were employed to represent the dimensions of the triple space (see Fig. 1). Unlike with traditional 6-sided dice, each number on these 8-sided dice was represented by a digit and not by a configuration of dots.

Figure 1: Eight-sided die.

In a second experimental condition, participants were asked to generate a triple, plot the three numbers on a simple grid and then draw a line connecting the points. On this grid the x axis coded the position of the three numbers in the sequence (first, second, third) and the y axis represented the dimension of the triple space, namely 1 to 8. Figure 2 illustrates how triple 2-4-6 would be represented on such a grid. In this manner, participants created a graphical representation of the pattern inherent in the number sequence. Such a graph visually exposed the linear relationship between adjacent numbers, thus eliminating the computational cost of inferring such a pattern from a numerical representation alone, a concept Scaife and Rogers (1996) refer to as *computational offloading*. A graphical representation conveys the slope inherent to triples that receive positive and negative feedback, as well as its acceleration. For example, a positively accelerating curve for triple 1-2-8 or a negatively accelerating curve for triple

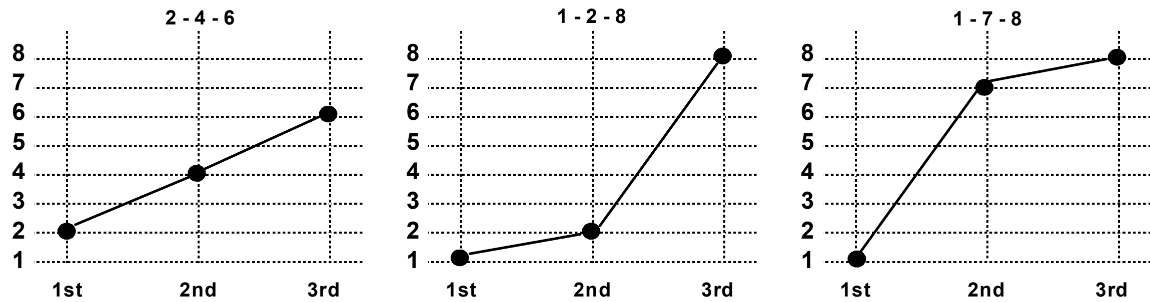

Figure 2: Plotted triples. The left-panel ('2-4-6' triple) was shown to the participants as part of the task instructions in the *Graph* condition.

1-7-8 (see Fig. 2). The inferential task in the *Graph* condition is thus isomorphic to the one in the control condition (and in the *Dice* condition). However, the *Graph* condition offers a perceptually compelling representation of the simple linear relationship between the consecutive numbers that compose a given triple. Such a representation may also direct the kinds of hypotheses formulated to describe such simple incremental patterns. In other words, examining plotted triples may 'limit abstractions' (Stenning & Oberlander, 1995).

Both the *Dice* and the *Graph* conditions enrich the distributed representation underpinning the hypothesis-testing process. In the *Dice* condition, the representation in enriched by the fact that reasoners can physically manipulate the three dimensions of the triple space in creating and selecting triples to test. In the *Graph* condition, the external representation created by plotting the triple affords 'perceptual inferences' (Larkin & Simon, 1987) that might facilitate the rule-discovery process.

Method

Design & Procedure

Participants were assigned, on a random basis to either the control condition or one of the two enriched representation conditions, *Dice* or *Graph.* Participants in all three conditions were tested individually in a quiet room. In the control condition, participants engaged in the 2-4-6 task in the absence of an enriched external representation. In the testing room participants sat down at a desk on which was placed an answer sheet headed with a paragraph of instructions followed by a table with two columns labelled *Number sequence* and *Feedback* and 18 unnumbered rows. On the first row the triple *2-4-6* and *YES* were printed in their respective column. The instructions read:

> The present task consists in discovering why certain numbers go together in a sequence. To start you off, I can tell you that 2-4-6 is a sequence that satisfies the rule I have in mind. In order to discover my rule, you should produce new sequences and for each sequence you produce I will tell you whether or nor it fits the pattern I am looking for. Number sequences can only be made up from numbers 1 to 8. You can produce as many or as few sequences as you wish, but proceed to tell me your best guess *only when you feel highly confident that you have discovered the rule that I have in mind*.

Participants wrote number sequences on the answer sheet and the experimenter entered the feedback in the adjacent column. Participants were not asked to formulate a hypothesis before the production of any triple nor offer any justification. When participants announced they were sufficiently confident to stop producing new triples and state their hypothesis they wrote their answer at the bottom of the answer sheet. The task ended after participants wrote down their answer; they did not continue with the task if they announced an incorrect hypothesis.

In the *Dice* isomorph participants sat a desk on which was placed three 8-sided dice and the same answer sheet. The dice displayed the numbers *2*, *4*, and *6*, in ascending order. Instructions were the same as in the control condition but for the third sentence that read:

> In order to discover my rule, you should produce new sequences using the three dice, each dice corresponding to one of the numbers in the sequence, and for each sequence you produce I will tell you whether or nor it fits the pattern I am looking for.

Participants generated new number sequences by manipulating the dice; they were not allowed to throw the dice to generate a triple. The experimenter transcribed each new triple on the answer sheet and provided feedback.

In the *Graph* condition instructions were on a separate sheet. They informed participants how to plot their triples on a small paper grid (4.5 cm by 6.5 cm) where the x axis was labelled *1st*, *2nd*, and *3rd* and the y axis ranged from 1 to 8. The instructions were the same as in the control condition except for the following modifications:

> In order to discover my rule, you should produce new sequences of three numbers. The new sequences can only be made up using numbers from 1 to 8. Once you have produced a new sequence, and before I tell you whether or not it fits the pattern I am looking for, I would like you to draw the number sequence on a grid, as illustrated below.

The grid shown to the subject was the triple 2-4-6 plotted as in the left panel of Figure 2. Instructions as to when to stop producing new sequences and announce a rule were the same as in the Control and *Dice* condition. The instructions

TABLE 1: Rule discovery performance in the *Dice* and *Graph* conditions (separate and combined) and in the control condition in terms of number of participants who announced the correct rule (and in percent), mean number of trials before announcements, mean number of triples that received positive and negative feedback, mean percentage of triples that received negative feedback, and mean homogeneity ratio (s.e. = standard error)

		Correct		Triples		Positive Triples		Negative Triples		Percent Negative		Homogeneity Ratio	
	N	n	%	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
Dice	30	16	53%	6.77	0.58	5.07	0.46	1.70	0.23	22.55%	3.00%	0.58	0.05
Graph	30	15	50%	6.47	0.73	4.23	0.40	2.23	0.53	28.33%	4.00%	0.52	0.05
COMBINED	**60**	**31**	**52%**	**6.61**	**0.47**	**4.65**	**0.31**	**1.97**	**0.28**	**26.87%**	**2.79%**	**0.55**	**0.04**
CONTROL	**61**	**19**	**31%**	**7.02**	**0.68**	**5.43**	**0.50**	**1.57**	**0.27**	**17.59%**	**2.33%**	**0.66**	**0.04**

next summed up the three-step procedure: (1) produce new sequence, (2) draw sequence on grid, (3) receive feedback. Plotted triples remained in full view of the participants as they produced new triples.

Measures

Performance in all three conditions was measured along six dimensions: (1) the proportion of participants who announced the correct *increasing sequence* rule; (2) the number of triples tested before announcing the rule; (3) the number of increasing or positive triples; (4) the number of nonincreasing or negative triples; (5) the proportion of tested triples that received negative feedback; (6) the homogeneity of the set of triples produced as gauged by the ratio of triples generated that increased by constant increments over all triples tested. If a, b, and c represent the first, second and third number in a triple, a constant increment triple is one where $(b - a) = (c - b)$. The higher this ratio, the narrower and more homogeneous the set of triples produced before announcing a rule.

Participants

One hundred and twenty one undergraduate students at Kingston University were recruited to participate in this study. Sixty one were assigned to the control group and 60 were assigned to one of two experimental conditions, 30 in the *Dice* isomorph and 30 in the *Graph* condition.

Results

Sixteen participants (or 53%) and 15 participants (or 50%) in the *Dice* and in the *Graph* conditions, respectively, announced the correct 'increasing sequence' rule. In contrast, 19 out of 61 (or 31%) announced the correct rule in the control condition. A chi-square analysis revealed that significantly more participants announced the correct rule in the experimental conditions (31 out 60) than in the control condition (19 out of 61), χ^2 (1) = 5.25 (a .05 rejection criterion was employed unless indicated otherwise).

Performance in the *Dice* and in the *Graph* conditions did not significantly differ along any of the six dimensions reported in Table 1, that is in terms of success rate (53% vs. 50%), triples tested before announcing the rule (6.77 vs. 6.47), number of triples that received positive (5.07 vs. 4.23) and negative (1.70 vs. 2.23) feedback, percentage of triples that received negative feedback (22.6% vs. 28.3%) and homogeneity ratio (0.58 vs. 0.52). In light of these results, the data from the two experimental conditions were combined and are reported as such in Table 1. Comparison with the control condition revealed that experimental participants produced a significantly greater percentage of negative triples (mean of 26.9%) than control participants (mean of 17.6%), $t(119) = 2.30$. The homogeneity ratio was significantly smaller for experimental participants (mean of 0.55) than for control participants (mean of 0.66), $t(119) = 2.03$. Rule discovery performance as measured in term of the triples produced before announcement and number of positive and negative triples did not differ significantly between the experimental and control conditions.

The rule discovery profile of participants who discovered the correct rule in the *Dice* and *Graph* conditions is reported in Table 2. The performance of these participants was statistically indistinguishable and as a result the data from both conditions were combined. Successful participants in the control condition tested, on average, 11.6 triples before announcing the correct rule whereas experimental participants tested 7.7 triples on average before announcing the correct rule, a significant difference, $t(48) = 3.22$. Control participants also tested significantly more triples that receive positive feedback than experimental participants (respective means of 8.5 vs. 5.4), $t(48) = 3.25$. The rule discovery profile of successful participants in the experimental and control conditions did not differ significantly in terms of negative triples tested, proportion of negative triples tested, and homogeneity ratio.

Discussion

In the study reported here the external component of a distributed representation of Wason's 2-4-6 inductive inference task was enriched in two different ways. In the *Dice* condition, the dimensions of the triple space were presented as three 8-sided dice that afforded physical manipulation. In the *Graph* condition, the simple linear

TABLE 2: Rule discovery performance by participants who announced the correct rule in the *Dice* and *Graph* condition (separate and combined) and in the control condition in terms of mean number of trials before announcements, mean number of triples that received positive and negative feedback, mean percentage of triples that received negative feedback, and mean homogeneity ratio (s.e. = standard error)

		Triples		Positive Triples		Negative Triples		Percent Negative		Homogeneity Ratio	
	N	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
Dice	16	7.75	0.75	5.50	0.72	2.25	0.23	30.55%	3.27%	0.47	0.06
Graph	15	7.73	1.02	5.33	0.64	2.40	0.50	29.87%	3.67%	0.46	0.06
COMBINED	**31**	**7.74**	**0.61**	**5.42**	**0.48**	**2.32**	**0.26**	**30.22%**	**2.41%**	**0.47**	**0.04**
CONTROL	**19**	**11.58**	**1.14**	**8.53**	**0.94**	**3.05**	**0.48**	**26.61%**	**3.36%**	**0.42**	**0.06**

relationship between numbers in each number sequence was plotted. It seems likely that these two manipulations influenced the triple production process differently, at least in the earlier stages of the task. However, the hypothesis-testing profiles they encouraged were statistically indistinguishable on all measured dimensions.

In turn, reasoners who engaged in the 1-8 version of the traditional 2-4-6 task (the control condition) were less likely to discover the rule with a success rate of 31%, compared to 53% and 50% for the *Dice* and *Graph* conditions, respectively. Diligence as gauged by the number of triples tested was as high in the control condition as in the enriched experimental conditions. However, in terms of relative composition, the set of triples produced by the control participants was more homogeneous (lower percentage of negative triples, higher homogeneity ratio), reflecting a less creative exploration of the space of triples.

The impact of a richer external representation was particularly noteworthy when the performance profile of successful participants was examined. Control participants needed to test on average 11.6 triples before announcing the correct rule whereas experimental participants needed to test on average 7.7 triples. Thus, in the absence of a richer external representation, the inferential challenge posed by the Wason task was considerably harder to overcome. The hard work put in by the successful control participants also explain the relatively high mean number of triples tested in the control condition overall.

The results reported here corroborate those of Vallée-Tourangeau and Krüsi Penney (in press). Across these two studies, the success rate in the different experimental conditions where the external representation was enriched ranged between 46% and 66% in contrast to the success rate in control, 'impoverished' conditions that ranged between 19% and 31%. Thus, configuring the external representation of either the dimensions of the triple space or of the linear relationships among numbers composing a triple, significantly improves reasoners' ability to infer the correct rule.

These results support the notion that a fruitful research program may result from the exploration of external or contextual determinants of successful rule discovery performance. This would represent, in our opinion, a more productive shift away from the goal of explaining successful rule discovery performance in terms of psychometric individual differences as endorsed by Wason (1960) himself and others (e.g., Baron, 2000; Vartanian et al., 2003). People obviously vary along many a psychometric dimension. There is hardly a less contentious fact in the history of psychology. Apart from deflecting attention away from the importance of external representation in reasoning and problem solving, a focus on individual differences implicitly endorses the representativeness of the hypothesis-testing profile recorded with the traditional Wason task. The current research programme is predicated, however, on the notion that there are no inherent qualities to the Wason original procedure that should make it such a privileged window onto hypothesis-testing behavior. Such orthodoxy in research practice and interpretation results from the very narrow conceptual and methodological path trodden by two generations of researchers since Wason originally published his findings.

Better understanding of representational factors that support sound reasoning informs and shapes pedagogical intervention (Lowrie & Kay, 2001), effective information dissemination (Kleinmuntz & Schkade, 1993), and the design of cognitive artefacts (Norman, 1993). For example, Sedlmeier and Gigerenzer (2001) demonstrated how sound Bayesian reasoning can be taught and better elicited using a graphical interface wherein probabilistic information is presented in terms of natural frequencies. In a similar vein, we have recently gathered data that illustrate how the quality of causal inferences drawn from covariation information depends critically on the presentation format of that information (Payton & Vallée-Tourangeau, in preparation). Uncovering representational effects in inductive reasoning tasks alerts researchers, designers, and teachers to the possibility of shaping sounder hypothesis-testing behaviour through an active manipulation of the external representation of the task facing reasoners.

Acknowledgment

This research was supported by a grant from the Economic and Social Research Council (UK) awarded to FVT.

References

Baron, J. (2000). *Thinking and deciding* (3rd Edition). New York: Cambridge University Press.

Cheng, P. C.-H. (1996). Scientific discovery with law encoding diagrams. *Creativity Research Journal*, *9*, 145-162.

Duncan, S. (1998). *Are there representational effects in a rule discovery task?* Unpublished manuscript, Bowling Green State University.

Gorman, M. E. (1995). Hypothesis testing. In S. E. Newstead and J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 217-240). Hover, East Sussex: Lawrence Erlbaum Associates.

Klayman, J. & Ha., Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211-228.

Kleinmuntz, D. N., & Schkade, D. A. (1993). Information displays and decision processes. *Psychological Science, 4*, 221-227.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-100.

Lowrie, T., & Kay, R. (2001). Relationship between visual and novisual solution methods and difficulty in elementary mathematics. *Journal of Educational Research*, *94*, 248-255.

Mahoney, M. J., & DeMonbreun, B. G. (1977). Psychology of the scientist: An analysis of problem-solving bias. *Cognitive Therapy and Research, 1*, 229-238.

Norman, D. A. (1993). *Things that make us smart*. Cambridge, MA: Perseus Books.

Payton, T., & Vallée-Tourangeau, F. (in preparation). *The impact of the format of covariation information on causal inferences*.

Poletiek, F. H. (2001). *Hypothesis testing behaviour*. Hove, East Sussex: Psychology Press Ltd.

Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, *45*, 185-213.

Schustack, M. W. (1988). Thinking about causality. In R. J. Sternberg and E. E. Smith (Eds.), *The psychology of human thought* (pp. 92-115). New York: Cambridge University Press.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130*, 380-400.

Simon, H. A. (1996). *The sciences of the artificial* (3rd Edition). Cambridge, MA: MIT Press.

Stanovich, K. E., & West, R. F. (2000). Individual difference in reasoning: Implications for the rationality debate? *Behavioral & Brain Sciences, 23*, 645-726.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: logic and implementation. *Cognitive Science*, *19*, 97-140.

Sutherland, S. (1992). *Irrationality, the enemy within*. London: Penguin.

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

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

Vallée-Tourangeau, F., & Krüsi Penney, A. (in press). The impact of external representation in a rule discovery task. *European Journal of Cognitive Psychology*.

Vallée-Tourangeau, F., & New, M. (1998). Rational hypothesis-testing strategies in a rule discovery task. In M. A. Gernsbacher and S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1078-1083). Mahwah, NJ: Lawrence Erlbaum Associates.

Vanderhenst, J.-B., Rossi, S., Schroyens, W. (2002). When participants are not misled they are not so bad after all: A pragmatic analysis of a rule discovery task. In W. D. Gray and C. D. Schunn (Eds.), *Proceedings of the Twenty Fourth Annual Conference of the Cognitive Science Society* (pp. 902-908). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129-140.

Wason, P. C. (1968). 'On the failure to eliminate hypotheses…' – a second look. In P. C. Wason and P. Johnson-Laird (Eds.), *Thinking and reasoning* (pp. 165-174). London: Penguin.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 21*, 179-217.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*, 87-122.