Title

Contrast-Class Cues and Dual Goal Facilitation in Wason's 2-4-6 Task: Evidence for an Extended Iterative Counterfactual Model

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

ISSN
1069-7977

Authors
Ball, Linden J.
Gale, Maggie

Publication Date
2005

Peer reviewed

Contrast-Class Cues and Dual Goal Facilitation in Wason's 2-4-6 Task: Evidence for an Extended Iterative Counterfactual Model

Maggie Gale (M.Gale@derby.ac.uk)
Institute of Behavioural Sciences, University of Derby
Western Road, Mickleover
Derby DE3 9GX, UK

Linden J. Ball (L.Ball@lancaster.ac.uk)
Department of Psychology, Lancaster University
Lancaster, LA1 4YF, UK

Abstract

Successful performance on Wason's (1960) 2-4-6 task is typically poor at approximately 20%. One reliable way to enhance solution rates is to use Tweney, Doherty, Worner, Pliske, Mynatt, Gross, and Arkkelin's (1980) logically-identical Dual Goal (DG) version of the task, where participants are requested to discover two complementary rules, one labeled 'Dax' (the standard 'ascending numbers' rule) and the other labeled 'Med' (i.e., any other number triple). Despite the robustness of the DG effect, the mechanism by which the DG paradigm facilitates performance has remained obscure. The present studies assessed various theoretical proposals by providing participants examples of Med triples of varying 'usefulness', as indexed by the cues that they provided for establishing a relevant contrast class to the Dax rule. Results showed that the usefulness of the Med exemplar had a significant effect on successful discovery of the Dax rule. We propose that the present DG results can best be accommodated by extending Oaksford and Chater's (1994) Iterative Counterfactual Model beyond its current focus on the standard single goal 2-4-6 task.

Keywords: Hypothesis testing; Wason's 2-4-6 task; dual-goal facilitation; contrast-class cues; iterative counterfactual model.

Introduction

The traditional paradigm for investigating hypothesis-testing behaviour is the 2-4-6 task. Introduced by Peter Wason in 1960, this deceptively simple task was designed to examine whether people conformed to the contemporary scientific philosophy of falsification (Popper, 1959). The value of falsification as a hypothesis testing strategy lies in the fact that any number of confirming instances cannot prove a hypothesis to be correct, whereas a single disconfirming instance is sufficient to disprove it. Thus, a strong test of any hypothesis is one that attempts its disconfirmation.

In the standard, single goal (SG) form of Wason's 2-4-6 task participants are asked to discover a rule for generating groups of three numbers (triples). They are given an example triple of 2-4-6 as one that conforms to the experimenter's rule, and they are asked to generate further triples that the experimenter classifies as either conforming or not conforming to the target rule. The participant continues to generate triples until they are confident that they know the rule, at which point they announce it. The to-be-discovered rule is 'any ascending sequence'. Despite the simplicity of the task only 20% of participants successfully announce the rule on their first attempt (e.g., Tukey, 1986; Wason, 1960).

Unsuccessful announcements tend to be restricted forms of the target rule such as 'numbers ascending by equal intervals' (Kareev, Halberstadt, & Shafir, 1993). Wason (1960) also noted that solvers and non-solvers could be differentiated in terms of both the quantity and quality of triples produced before rule announcement: Solvers generated reliably more triples of which a greater proportion received negative feedback. Wason labeled this tendency of non-solvers to test triples that conformed to their hypothesis as a 'verification bias' (more commonly referred to nowadays as a 'confirmation bias').

Early criticism of the task came from Wetherick (1962), who argued that it is the initial 2-4-6 exemplar that induces participants to form the overly restricted hypothesis. Van der Henst, Rossi, and Schroyens (2002) provided support for this view. They invoked relevance theory, arguing that participants would expect that information provided by the exemplar triple was relevant to the task, which would lead to an initial hypothesis that was restricted by the information contained in the exemplar. An experimental manipulation that caused participants to believe that the regularities in the exemplar were generated by chance did, indeed, undermine the usual tendency for people to produce overly restrictive hypotheses.

Recent research by Cherubini, Castelvecchio, and Cherubini (2005) has clarified the process of initial hypothesis generation. In a series of ingenious experiments they varied the type and number of perceivable relationships in the example triples (e.g., ascending numbers, equal intervals, even numbers etc.). Results showed that participants preserved much more information in their initial hypotheses when the presented exemplars contained 'high information' regularities rather than 'low information' regularities. For example, given the triples 2-4-6 and 3-5-7 people prefer the hypothesis 'numbers ascending by two' rather than 'ascending numbers' because the former hypothesis conveys more information. The studies of Van der Henst et al., and Cherubini et al., make important contributions to elucidating how initial hypotheses are formed. However, both sets of researchers are silent on the question of *why* participants fail to test their hypotheses according to the Popperian ideal of falsification and thus discover the correct rule.

Dual Goal Instructions

A key issue of note in the history of the 2-4-6 task is that although reported rates of success are typically poor for the task in its standard form (at around 20%), one reliable way to improve success rates is to use Dual Goal (DG) instructions as introduced by Tweney, Doherty, Worner, Pliske, Mynatt, Gross, and Arkkelin, (1980). In this logically and formally identical version of the task, instructions are varied such that participants are asked to discover two complementary rules, one labeled 'Dax' the other 'Med'. Participants are given 2-4-6 as an example of the 'Dax' rule and are then asked to generate triples that are classified as conforming to either the 'Dax' rule (which produces 'ascending sequences') or the 'Med' rule (which produces all other sequences). This manipulation causes task success to rise dramatically, with rates of over 60% commonly being reported (e.g., Farris & Revlin 1989a, 1989b; Tukey, 1986; Wharton, Cheng, & Wickens, 1993). Despite the robustness of this DG effect there is little consensus as how best to account for this phenomenon, although a variety of theories have been proposed. We turn now to examining these theories and their predictions.

Goal Complementarity Theory: According to the goal complementarity theory, DG facilitation is caused by the combination of a 'positive test strategy' (Klayman and Ha, 1987) and the complementary nature of the Dax and Med rules. It is proposed (e.g., Wharton et al., 1993) that participants test the Dax hypothesis with positive triples until satisfied with their formulation of the rule (e.g., 'numbers rising with equal intervals'), at which point they turn their attention to the Med rule. Positive testing of the current Med hypothesis (e.g., 'numbers rising with unequal intervals') leads to unexpected Dax feedback and engenders the need to reformulate the Dax rule. A prediction of the goal complementarity theory is, therefore, that it is the production of triples rising by non-equal increments (so-called 'variable positives') to test the Med hypothesis that leads to the progressive refinement of the Dax hypothesis and eventual task success. A strong version of the goal complementarity theory, however, was recently undermined in a study by Gale and Ball (2003), who manipulated the relationship between the Dax and Med rules such that they were no longer complementary, yet they still found a large facilitatory effect of DG instructions (see also Vallée-Tourangeau, Austin, & Rankin, 1995, for related evidence).

Information Quantity Theory: Proposed by Wharton et al. (1993), information quantity theory is an intuitively appealing proposal that invokes the observations that: (1) solvers produce more triples before rule announcement than non-solvers, and (2) DG instructions typically lead to higher levels of triple production. Evidence for this proposal has been mixed. Klayman and Ha (1989) examined participants' 'best guesses' after they had generated 6, 12 and 18 triples, and showed that increasing numbers of participants had correctly guessed the target rule at each stage. By contrast, Vallée-Tourangeau et al. (1995) and Gale and Ball (2003) showed that DG instructions facilitate performance even when the number of triples generated is held constant across SG and DG paradigms.

Triple Heterogeneity Theory: Triple heterogeneity theory (Vallée-Tourangeau et al., 1993) appeals to the idea that successful solvers produce not only a greater number of triples than non-solvers, but also that generated triples are of a greater variety, reflecting a wider exploration of the hypothesis space. To test this idea Vallée-Tourangeau et al. (1993) introduced a simple method of codifying triples as either 'positive' (conforming to the target rule) or 'negative' (not conforming to the target rule). Positive triples were further subdivided into 'constant positives' (rising by equal intervals) or 'variable positives' (rising by non-equal intervals). Negative triples were categorized into eight possible 'negtype' forms: Equal numbers, descending numbers, two equal numbers and one larger number etc. Vallée-Tourangeau et al. reported results of a study demonstrating increases in the variety of triples people generated (e.g., more negtypes and posvars) contingent upon the presence of DG instructions. We are concerned, however, that this account is largely descriptive in nature such that it does not offer any detailed explanation of the *mechanism* by which DG instructions promote a wider exploration of the problem space.

Contrast Class Theory: Gale and Ball (in press) conducted a close analysis of the triples generated by successful and non-successful participants in SG and DG conditions and noted that the production of at least one 'descending' triple was reliably associated with task success. Gale and Ball argue that the production of a descending triple provides a salient 'contrast class' that causes the participant to focus on the 'ascending' aspect of the Dax hypothesis, thus promoting discovery of the target Dax rule. Further analysis showed that participants given DG instructions were far more likely to produce a descending triple than those given SG instructions.

Aims of the Study

The primary aim of the present study was to explore further Gale and Ball's (in press) contrast class theory by providing participants not only with the standard 2-4-6 Dax exemplar but also a Med exemplar that provided either a 'useful' or a 'non-useful' contrast class cue. Half of our participants were, therefore, told that an example of the Med rule was the triple 6-4-2 (useful contrast class information). We predicted that this Med exemplar would promote identification of the Dax rule since 6-4-2 and 2-4-6 are oppositional on the salient – and crucially *relevant* – dimension of 'ascending' versus 'descending'. The other half of our participants were told that an example of the Med rule was the triple 4-4-4 (non-useful contrast class information). We predicted that this Med triple would *not* promote successful Dax discovery as 4-4-4 and 2-4-6 are oppositional on the salient – but *non-relevant* – dimension of 'three identical numbers' versus 'three different numbers'. Our experiment also allowed us to assess other theories of DG facilitation. For example, by requiring all participants to generate exactly 10 triples we were able to control for information quantity whilst also being able to analyze the 10 triples produced so as to quantify key aspects of triple heterogeneity.

Method

Participants

Forty-three, first-year psychology students from the University of Derby took part in the study on a voluntary basis. None had received any teaching relating to reasoning or logic.

Design

An independent-measures design was employed with the manipulation reflecting the usefulness of the contrast class cue (CCC) that was salient in the presented example of the Med rule. One group of participants received a useful CCC (6-4-2) and the other group were given a non-useful CCC (4-4-4). Participants were randomly assigned to the two conditions. The DG paradigm was used in both conditions of this experiment.

Procedure

Participants were tested individually in a quiet laboratory. Standardized DG instructions were read out to all participants as follows: ***"****I have in mind two rules that specify how to make up sequences of three numbers (triples), and your task is to discover these rules. Triples that fit one of my rules are called Dax triples and those that fit my other rule are called Med triples…"*. All participants were given 2-4-6 as an example Dax triple. Those in the useful CCC condition were given 6-4-2 as a Med exemplar, while those in the non-useful CCC condition were given 4-4-4 as a Med exemplar. Participants were then asked to produce exactly 10 triples, and they received feedback for each triple in the form of 'Dax' or 'Med'. After 10 triples had been generated participants were asked to write down their best guess at the two rules. In line with the procedure of Gorman (1992), participants were allowed only one guess at these rules.

Results

Solution Success Across Conditions

Analysis of the solution-success data was concerned only with correct announcements relating to the Dax rule. This scoring method is standard practice in studies with DG task variants (cf. Vallée-Tourangeau et al., 1993). Table 1 shows the frequency of correct rule announcements for the Dax rule in each of the experimental conditions. It is clear that the usefulness of the CCC that had been provided had a dramatic effect on success rates for this task: 74% of participants who received a useful CCC made a correct rule announcement, while only 20% of participants who received a non-useful CCC were successful on the task. A chi-square analysis showed this effect to be highly significant, $\chi^2(1) = 12.44$, $p < .001$. Interestingly, these success rates are very similar to those typically reported in the literature for the DG and SG paradigms respectively (e.g., Wharton et al., 1993). This observation suggests that it may not be DG instructions per se that lead to task success in the DG paradigm, but rather that DG instructions facilitate participants' production of a salient contrast class which, in turn, promotes successful discovery of the target Dax rule.

Table 1: Frequency of correct DAX rule announcements by condition.

		Success	
Condition	N	Solver	Non-solver
Useful CCC	23	17	6
Non-Useful CCC	20	4	16

Production of Descending Triples

The purpose of this study was to test the idea that the key to success on the DG 2-4-6 task relates to the availability or discovery of useful contrast class information that facilitates identification of the potential scope of the Dax rule. In our previous research (e.g., Gale & Ball, in press) we have noted that successful solvers are those who uncover at least one descending triple during their hypothesis testing, and that DG instructions promote the production of at least one such descending triple.

With these findings in mind, we decided to examine the present dataset for any effect of the presence versus absence of the participant's production of a descending triple on their task success. Table 2 presents data collapsed across the useful versus non-useful CCC manipulation. The table shows that while all but 10 participants produced at least one descending triple, there was no single instance of a participant solving the task in the absence of a descending triple. A chi-square analysis indicated that this effect was highly reliable, $\chi^2(1) = 12.44$, $p < .001$.

Table 2: Frequency of correct Dax announcements by presence vs. absence of at least one descending triple.

		Success	
Descending Triple	N	Solver	Non-solver
Absent	10	0	10
Present	33	21	12

The data were further explored to ensure that the type of CCC that had been provided had a reliable influence on whether or not at least one descending triple was produced (see Table 3). This analysis (collapsing across solver vs. non-solver) showed that all participants receiving a useful CCC produced at least one descending triple, while only 50% of participants receiving a non-useful CCC produced a descending triple, $\chi^2(1) = 14.99$, $p < .001$. It can thus be concluded that the CCC manipulation was successful in terms of its capacity to induce the generation of triples of a particular type (i.e., descending numbers).

Table 3: Frequency of production of at least one descending triple by condition.

		Descending Triple	
Condition	N	Present	Absent
Useful CCC	23	0	23
Non-Useful CCC	20	10	10

Triple Type

One aim of the present study was to explore other accounts of DG facilitation such as Vallée-Tourangeau et al.'s (1993) triple heterogeneity theory. Triples were, therefore, codified using the system introduced by Vallée-Tourangeau et al. (see above) and statistical comparison were pursued between the triple types generated by solvers and non-solvers (Table 4). The only significant difference on any of these measures of triple variety related to the number of descending triples that were produced, with solvers producing reliably more than non-solvers, $t(41) = 5.77$, $p < .001$.

One further measure of triple heterogeneity was also calculated: The *overall number* of different types of triple generated, whether positive or negative. Again there was no significant difference across task success. These results provide further evidence against both the triple heterogeneity account and the goal complementarity account of facilitated performance on the DG task. The triple heterogeneity theory proposes a wider exploration of the problem space, as indexed by the variety of triples generated; the present analyses suggest that variety of triples does not, in fact, influence success on the DG task. The goal complementarity theory suggests that it is the production of variable positive triples that should be a key feature of successful DG solving. Again, the data presented here do not support this notion.

Table 4: Mean number of triple types produced by success, with standard deviations in parentheses.

	Success	
Heterogeneity measure	Solver	Non-solver
Constant Positives	4.71 (1.47)	3.86 (2.10)
Posvars	0.71 (1.06)	0.64 (1.78)
Negtypes	1.62 (0.97)	2.41 (2.36)
Negative feedback	4.62 (0.97)	5.18 (1.74)
Descending triples	3.81 (1.33)	1.36 (1.73)
Variety	7.04 (1.20)	6.90 (1.63)

Discussion

The study aimed to explore accounts of facilitated performance on DG versions of Wason's 2-4-6 task, with particular attention being given to Gale and Ball's (in press) contrast class ideas. Data deriving from the study support contrast class predictions, while offering little support for either the goal complementarity or triple heterogeneity theories. By holding the number of triples required constant we were also able to show that the information quantity account does not provide an adequate explanation for the different success rates seen in our experimental conditions.

How, then, can we explain the detailed theoretical underpinnings of the contrast class account? We appeal here to what we believe to be the most psychologically plausible account of behaviour on the SG version of the 2-4-6 task, that is Oaksford and Chater's (1994) Iterative Counterfactual Model (ICM), itself a development of Farris and Revlin's (1989a, 1989b) earlier counterfactual strategy. The ICM focuses on how hypotheses are created rather than on how they are tested, with one key aspect being how hypotheses are revised when falsifying evidence is obtained.

The operation of the ICM can best be illustrated with an example: Imagine a participant's working hypothesis, H, is 'even numbers ascending by two'. According to the ICM, the participant then generates an alternative hypothesis, H', that is complementary to H for a single property (a procedure that reflects Tschigi's, 1980, 'vary one thing at a time' hypothesis-testing strategy). Thus, given the exemplar 2-4-6 and the H of 'even numbers ascending by 2', an H' of 'odd numbers ascending by two' could be generated. A positive example of H' (e.g., 3-5-7) would be generated and tested. If this test proves positive then both H and H' must be false, and a new H must therefore be generated based on the common properties of the triples generated so far. In the example given, a new H of 'number ascending by two' might be generated. The process is then repeated with a new H' (e.g., 'numbers descending by two') being posited and tested. The cycle continues, with each new H being informed by the common properties of the pool of tested triples that have so far received positive feedback, until a point is reached where the current H satisfies the reasoner and the hypothesis is announced as the target H.

At its inception, Oaksford and Chater's (1994) ICM was concerned with hypothesis generation in the SG task, and not with explaining facilitated performance in the DG paradigm. We believe, however, that the ICM can be extended to encompass DG facilitation (cf. Gale & Ball, in press). We propose, for example, that reasoners are not only sensitive to the information regularities in Dax and Med triples (cf. Cherubini et al., 2005) but also to the contrast class that such regularities invoke. Thus, at the second iteration in our example above, the participant would be given the feedback that '6-4-2 is a Med triple'. H and H' are complimentary for the property ascending versus descending, and, we suggest, it is this opposition which establishes a salient contrast class that promotes the insight that ascending and descending are the relevant characteristics of the pool of triples tested.

Supporting evidence for an extended ICM that incorporates ideas of contrast-class identification is provided by the solution-success results of the present study. In relation to the non-useful CCC condition, for example, we had expected that participants who were given the illustrative 4-4-4 Med triple would generate a 'three equal numbers' hypothesis for Med, and would, therefore, be lured toward considering a (non-relevant) 'three different numbers' hypothesis for Dax. Our results indeed showed instances of participants going as far as announcing a final Dax rule as being 'three different numbers' in the non-useful Med condition. This supports the notion that reasoners are highly susceptible to focusing on the apparent relevance of available contrast class cues during their hypothesis testing.

Follow-Up Study

Given the previous interpretation of our results we decided to run a follow-up study involving a DG condition that presented an example Med triple that provided non-useful contrast class information, but in a subtly different way to the 4-4-4 exemplar in the previous study. To this end we presented 9-8-1 as an illustrative Med triple. Whereas 4-4-4 and 2-4-6 contrast on the single, non-relevant dimension of 'three identical numbers' versus 'three different numbers', 9-8-1 and 2-4-6 contrast on: (1) several, non-relevant dimensions (e.g., 'mixed odd and even numbers' vs. 'only even numbers', 'unequal intervals' vs. 'equal intervals', and 'middle number is not arithmetic mean of the outer numbers' vs. 'middle number is the arithmetic mean of the outer numbers'); and (2) a single, relevant dimension (i.e., 'descending numbers' versus 'ascending numbers'). We predicted that the multiplicity of the available contrast class cues would, at the very least, render performance on this 9-8-1 variant of the DG task as impoverished as that seen with the 4-4-4 variant previously described.

Method

Twenty-two University of Derby students took part in the follow-up study. The method employed was exactly the same as that for the previous study, with the exception that the Med example provided was 9-8-1.

Results

Eight of the 22 participants (36%) solved the 9-8-1 DG task. A chi-square analysis was employed to compare the performance of these participants with those in the original study. This indicated a reliable difference in success rates across conditions, $\chi^2(2) = 13.5$, $p = .001$. Examination of adjusted residual values showed that the difference was between those given 6-4-2 as a Med exemplar and those given either 4-4-4 or 9-8-1. This observation supports our proposals for an extended ICM to capture successful hypothesis testing in DG versions of the 2-4-6 task.

Conclusion

The present studies offer insights into how Oaksford and Chater's (1994) Iterative Counterfactual Model may be extended to account for the facilitatory effect of DG instructions on Wason's (1960) 2-4-6 task. In particular, we propose that the inclusion of a mechanism within the ICM that directly promotes the identification and use of salient contrast class cues relating to Dax and Med triples can provide psychologically plausible explanations of a number of important DG phenomena (cf. Gale & Ball, in press).

First, such a mechanism accounts for why the production of descending triples is seen to mediate between DG instructions and discovery of the Dax rule, as descending triples, categorized by the experimenter as instances of the Med rule, will help establish an appropriate contrast class (descending vs. ascending) that captures the true scope of the Dax rule. Second, this mechanism accounts for why (as seen in the present studies) different Med exemplars either promote or inhibit discovery of the target Dax rule. This is because some Med exemplars (e.g., 4-4-4 or 9-8-1) will provide non-useful contrast class cues that are of little or no help in Dax identification, whilst other Med exemplars (e.g., 6-4-2) readily cue the identification of a highly useful Dax-Med contrast class (i.e., descending vs. ascending).

In sum, Oaksford and Chater's (1994) Iterative Counterfactual Model appears to provide a compelling foundation for explicating the detailed processes involved in hypothesis generation and testing within both SG and DG 2-4-6 tasks. We acknowledge, however – as, indeed, do Oaksford and Chater (1994) – that the ICM needs further refinement and evaluation to provide a fully coherent account of all 2-4-6 phenomena. At some point we are certain that researchers will have to grapple once again with the complexities of direct recording, coding and analysis of participants' moment-by-moment hypothesis formulation and triple generation in order to track more clearly the complex relationship that seems to exist between hypotheses, triple instances, feedback and rule discovery in Wason's 2-4-6 paradigm.

References

Cherubini, P., Castelvecchio, E., & Cherubini A. M. (2005). Generation of hypotheses in Wason's 2-4-6-task: An information theory approach. *Quarterly Journal of Experimental Psychology*, *58A,* 309-332.

Farris, H., & Revlin, R. (1989a). Sensible reasoning in two tasks: Rule discovery and hypothesis evaluation. *Memory and Cognition, 17,* 221-232.

Farris, H., & Revlin, R. (1989b). The discovery process: A counterfactual strategy. *Social Studies of Science, 19,* 497-513.

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

Gale, M., & Ball, L. J. (in press). Dual-Goal facilitation in Wason's 2-4-6 task: What mediates successful rule discovery? *Quarterly Journal of Experimental Psychology.*

Gorman, M. E. (1992). Experimental simulations of falsification. In M. T. Keane, & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking: Vol. 1*. Hemel Hempstead: Harvester Wheatsheaf.

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

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

Oaksford, M., & Chater, N. (1994). Another look at eliminative and enumerative behaviour in a conceptual task. *European Journal of Cognitive Psychology, 6,* 149-169.

Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.

Tschigi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51,* 1-10

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

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.

Van der Henst, J-P, 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, & C. D. Schunn (Eds.) *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 902-907). Mahwah, NJ: LEA.

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

Wetherick, N. E. (1962). Eliminative and enumerative behaviour in a conceptual task. *Quarterly Journal of Experimental Psychology, 14,* 129-140.

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