Available online at www.sciencedirect.com
ScienceDirect

Brain and Cognition 62 (2006) 74–79

Brain and Cognition

www.elsevier.com/locate/b&c

Learned irrelevance and response perseveration in a total change dimensional shift task

J.H.R. Maes *, J. Vich, P.A.T.M. Eling

NICI/Biological Psychology, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands

Accepted 30 March 2006
Available online 8 May 2006

Abstract

Thirty-six healthy participants received a discrimination learning task requiring the identification of a relevant stimulus dimension. After successful learning, the relevant dimension was shifted unannounced. All exemplars of the two dimensions presented after the shift were novel, implying a 'total change' design. In three experimental conditions, participants could either make only errors reflecting perseveration of responding to the former relevant dimension, continued ignoring of the former irrelevant dimension, or both. After the shift, the participants in the perseveration condition made fewer errors than did those in the other two conditions, which did not differ. These results imply a predominance of the learned irrelevance mechanism even when any direct transfer of learning about exemplars in the pre-shift phase is precluded.

Keywords: Associative learning; Dimensional shift; Discrimination learning; Learned irrelevance; Response perseveration; Stimulus dimension

1. Introduction

A number of neuropsychological tests imply a shift in the rule that enables correct responding, such as the Wisconsin Card Sorting Test (WCST). These tests are frequently used to assess executive functions in a variety of clinical populations, such as people suffering from frontal lobe dysfunction, Parkinson's disease, Huntington's disease, or schizophrenia. These populations often display impaired task performance (see Lawrence, Sahakian, Rogers, Hodges, & Robbins, 1999; Pantelis et al., 2004; Ravizza & Ciranni, 2002, for a few recent examples), which may reflect a deficit in one or more cognitive capacities. Cognitive functions proposed to underlie executive functioning in general are often described very broadly, such as in terms of memory, planning, motor control, and attention, and there is a great need for a more detailed characterization of mechanisms underlying each of these functions. Fortunately, a number of previous studies already explicitly addressed some mechanisms potentially involved in rule-shift tasks, such as the ability to generate or maintain rules, the ability to no longer respond to a previously valid, but currently invalid rule, and the capacity to direct attention to a previous irrelevant rule (e.g., Barceló & Knight, 2002; Gauntlett-Gilbert, Roberts, & Brown, 1999; Joosten, Coenders, & Eling, 1995; Lanser, Berger, Ellenbroek, Cools, & Zitman, 2002; Lawrence et al., 1999; Li, 2004; Owen et al., 1993).

A previous study (Maes, Damen, & Eling, 2004), which was based on experiments reported by Gauntlett-Gilbert et al. (1999), Lawrence et al. (1999), and Owen et al. (1993), focussed on two such mechanisms: response perseveration and learned irrelevance. Response perseveration refers to the tendency to continue to respond to a particular stimulus or stimulus dimension, even in the face of negative feedback. Learned irrelevance refers to reduced attention directed at a stimulus or a stimulus dimension that proves to be task irrelevant. The standard WCST, or any discrimination learning task that is conceptually similar to this test, does not allow to differentiate between the role of

---

[*] Corresponding author. Fax: +31 24 3616066.
E-mail address: r.maes@nici.ru.nl (J.H.R. Maes).

0278-2626/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandc.2006.03.010

perseveration and learned irrelevance to post-shift responding. Difficulties in learning the novel, post-shift discrimination may reflect continued responding to the former correct dimension and exemplars, continued inattention to the former irrelevant (but now relevant) dimension and exemplars, or both. However, using appropriate experimental procedures, it is possible to disentangle their contribution. Specifically, a condition in which the former irrelevant dimension is effectively eliminated by holding it constant (e.g., if 'colour' is the irrelevant pre-shift dimension, all exemplars in the post-shift phase are presented in the same colour), eliminates the possibility that learned irrelevance affects task performance. However, post-shift performance may still be affected by the tendency to keep responding to the former relevant dimension, which is still present. Similarly, if the previously relevant dimension is eliminated, the response perseveration mechanism no longer can play a role, thereby enabling the unconfounded assessment of the learned irrelevance mechanism.

Using these procedures, Maes et al. (2004) found a larger contribution of the learned irrelevance than the perseveration mechanism to the performance of healthy human participants. However, although the post-shift phase did imply a shift in relevant stimulus dimension (a so-called extradimensional shift) in this study, the exemplars of the dimension(s) of major interest (either the relevant or irrelevant, or both, depending on the experimental condition) were identical to those used for the corresponding dimensions in the pre-shift phase. This might be particularly conducive to occasion associative learning processes, which might underlie learned irrelevance and perseveration (see Maes et al., 2004) and which are commonly held to operate on exemplars rather than on abstract dimensions. In fact, the study was explicitly designed in this manner to maximize the chance of occurrence of each of the two proposed processes.

A serious drawback of this approach is that it is not possible to unequivocally claim, on the basis of the obtained results, that the proposed mechanisms operate at the level of dimensions. This is particularly problematic given that WCST (-like) tasks are (implicitly) assumed to invoke processes related to the use of dimensions. Therefore, the generality of these earlier results and the comparability with at least some previous (clinical) studies may be limited. To demonstrate that learned irrelevance and perseveration affect the use of specific dimensions, it is necessary to employ a so-called 'total-change' design (Slamecka, 1968), in which all exemplars in the post-shift phase are novel. This feature precludes any effect on post-shift task performance of learning relatively simple exemplar-response associations, or learning to ignore specific exemplars, that is based on performing the pre-shift task. Indeed, it could be argued that, if participants only pay attention to exemplars, an associative learning approach predicts no difference between the various experimental conditions as used in Maes et al. (2004) when adopting a total change procedure if learned irrelevance and perseveration are processes that exclusively operate at the level of exemplars. Instead, if participants primarily approach the task in terms of stimulus dimensions it remains an empirical question whether learned irrelevance and perseveration mechanisms will operate and, if so, which of these mechanisms will play the most dominant role.

The results of three previous studies that assessed the separate contribution of learned irrelevance and perseveration in healthy subjects (next to patients), and that adopted a total change design, are mixed. Gauntlett-Gilbert et al. (1999) seemed to find more errors based on learned irrelevance than perseveration, whereas Owen et al. (1993) seemed to have found just the opposite, but no appropriate statistical analyses were reported regarding these differences in these two studies. Finally, Lawrence et al. (1999) did not find a significant difference in frequency between the two error types. Moreover, two of these studies failed to adopt a fully counterbalanced design with respect to the concrete dimensions used in the different phases of the experiment. This makes it impossible to draw any firm conclusions because one type of shift (e.g., from colour to shape) may be more difficult or more easy than another type of shift (e.g., from colour to number). Therefore, the purpose of the present experiment was to explicitly assess which process, learned irrelevance or perseveration, if any, plays the most dominant role when using a fully counterbalanced, total change design. Next to providing necessary information on the generality of the previous findings, the outcome of this study also speaks to the issue of the level (dimensions or exemplars) at which the proposed mechanisms exert their effects in dimensional shift tasks in general.

2. Materials and methods

2.1. Subjects

Forty-six participants, 50% male, volunteered to participate. The subjects consisted of both students and non-students and were all inhabitants of the city of Nijmegen, The Netherlands, or surrounding communities. Ten participants did not successfully complete the pre-shift phase within the maximum number of 80 trials (see Procedure below). The data from these subjects were excluded from data analysis because we were explicitly interested in the number of errors in the post-shift phase for subjects that had fulfilled the pre-shift learning criterion (see below). The remaining participants were 17 males and 19 females with a mean age of 30 years (range: 19–56). The three experimental groups ($n = 12$) were roughly equivalent with respect to sex ratio, mean age, and education.

2.2. Apparatus and stimuli

The experiment was run on a laptop with a 15 in. monitor, using an experiment-generator program. The participants were individually tested in their home environment in a quiet room. Each trial consisted of the presentation of a pair of stimuli, one stimulus presented on the left and one on the right side of the computer screen. Each stimulus could fill a virtual square of maximally 10 × 10 cm and consisted of one or more geometrical figures. Specifically, the stimuli could vary on two of three dimensions: colour, shape, and number. The exemplars used within each dimension and experimental phase (see below for further details) are shown in Table 1. That table also shows the exemplar used when a particular dimension was not varied (i.e., held constant) within each phase, and the exemplars to be chosen if a particular dimension was the relevant dimension.

For example, if shape was the relevant dimension and colour was the irrelevant dimension in the pre-shift phase, the number dimension was held constant and all stimuli consisted of three figures. Moreover, choosing the circle or the rhombus figure was a correct response, and choosing the triangle or trapezium was an incorrect response. In each experimental phase, each pair of stimuli consisted of a stimulus containing a correct exemplar of the relevant dimension and a stimulus containing an incorrect exemplar of the relevant dimension. Moreover, one of the stimuli of each pair contained a (dummy) correct exemplar of the irrelevant dimension and a (dummy) incorrect exemplar of the irrelevant dimension. The correct exemplar of the relevant dimension was presented concurrently with the (dummy) correct exemplar of the irrelevant dimension on one half of the trials, and concurrently with the (dummy) incorrect exemplar of the irrelevant dimension on the other half. For example, if shape was relevant and colour irrelevant, a pair of stimuli could consist of three purple circles (correct stimulus) and three light-green triangles (incorrect stimulus), or of three light-green circles (correct) and three purple triangles (incorrect), etc. The spatial location of the correct stimulus was left on one half of the trials and right on the other half (randomly determined). Fig. 1 shows some examples of stimulus pairs used.

Each pair of stimuli was presented until the subject pressed either the ‘1’ or the ‘2’ key on the computer keyboard, indicating a choice of the left or right stimulus, respectively. Feedback about choice accuracy was presented immediately after the subject’s response and consisted of the 1.5-s presentation of either the word ‘Correct’ written in blue or of the word ‘Incorrect’ written in red. Feedback was presented in the upper part of a white screen. The next stimulus pair was presented immediately after the feedback.

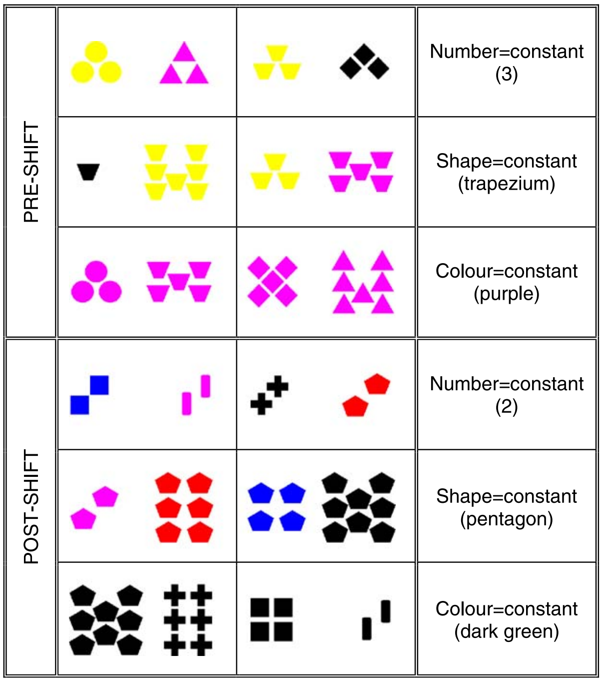

Fig. 1. Examples of stimulus pairs presented on the choice trials.

Table 1
Dimensions and exemplars used in the pre- and post-shift phases

	Shape	Colour	Number
Pre-shift	Circle*	Light green*	1*
	Rhombus*	Black*	5*
	Triangle	Yellow	7
	Trapezium	Purple	3
Post-shift	Cross*	Blue*	4*
	Square*	Red*	6*
	Bar	Pink	8
	Pentagon	Dark Green	2

Note. Underlined exemplars were used if the corresponding dimension was held constant.
* Correct exemplar if the corresponding dimension was relevant.

2.3. Procedure

The participant was seated in front of the computer monitor and read the following instructions (translated from Dutch):

During the experiment, you will repeatedly see two figures. Each time, one of these figures is correct; the other is incorrect. There are rules that determine which figure is correct. It is your task to detect these rules as fast as possible. Press the ‘1’ key if you think that the figure on the left is correct. Press the ‘2’ key if you think that the figure on the right is correct. Try to answer as fast as possible. After each choice, the computer will give you feedback, telling you whether your choice was correct or incorrect. The rules will change during the experiment.

The experiment started after the experimenter had made sure that the participant understood the instructions. All participants first received a maximum of 80 trials to sort out the correct dimension and exemplars (pre-shift phase). The pre-shift task was considered to be solved if the participant had made eight consecutive correct choices. Immediately after the eighth correct choice, the post-shift phase was initiated without notice. Again, the participant had to find the correct dimension and exemplars. Although the expression used in the instruction was in plural (‘the rules will change’) there was only one shift. This phase (and the experiment) ended upon making eight consecutive correct

Table 2
Outline of the experiment

Condition	Pre-shift		Post-shift	
	Relevant dimension	Irrelevant dimension	Relevant dimension	Irrelevant dimension
P	Dimension 1	Dimension 2	Dimension 3	Dimension 1
LI	Dimension 1	Dimension 2	Dimension 2	Dimension 3
P+LI	Dimension 1	Dimension 2	Dimension 2	Dimension 1

Note. The identity of dimensions 1, 2, and 3 (shape, number, or colour) were fully counterbalanced between and within conditions. The exemplars from the common dimensions used in the post-shift phase were different from those used in the pre-shift phase.

choices or after 80 trials, whichever came first. The participants were assigned to one of three experimental conditions that differed in the nature of the post-shift phase. Table 2 displays an outline of the experimental design.

Specifically, for the subjects in Condition P ('perseveration'), the dimension that had been relevant in the pre-shift phase became the irrelevant dimension in the post-shift phase. The dimension that was held constant in the pre-shift phase became the relevant dimension in the post-shift phase. Therefore, this condition allowed perseverative errors (continued choice of the former relevant dimension) but not errors due to learned irrelevance because the former irrelevant dimension was no longer present. In Condition LI ('learned irrelevance'), the pre-shift irrelevant dimension became the post-shift relevant dimension and the constant dimension from the pre-shift phase became the irrelevant dimension in the post-shift phase. This ensured the possibility to make errors based on learned irrelevance (continued neglect of the former irrelevant dimension) but no errors based on response perseveration. Finally, in Condition P+LI, the pre-shift relevant dimension became the post-shift irrelevant dimension and the pre-shift irrelevant dimension became the post-shift relevant dimension. This enabled the occurrence of both types of error.

There were six versions of each condition, with each version having a unique combination of concrete identity of pre-shift relevant and irrelevant dimensions. The identity of the post-shift relevant and irrelevant dimensions within each version was determined by the nature of the post-shift task that had to be created within the corresponding condition. Each version within each condition was performed by two subjects, implying a full counterbalanced design, within and between conditions, with respect to the specific nature of the different types of dimension.

2.4. Dependent measure and statistical analysis

The main dependent measure examined was the number of errors made in the pre- and post-shift phases. (Analyses using the number of pre- and post-shift trials revealed similar results). Error frequencies were subjected to analyses of variance (ANOVAs). However, as a particularly sensitive test, we were especially interested in the outcome of a Helmert contrast, which was explicitly based on our previous results (Maes et al., 2004). Specifically, these results implied a closer correspondence with respect to post-shift error scores between conditions LI and P+LI than between conditions P and P+LI, suggesting that performance in the P+LI condition was mainly determined by the learned irrelevance mechanism. In fact, in Experiment 1 of that study, performance in Condition P did not significantly differ from that in two control conditions. A Helmert contrast compares each level of a factor with the mean of subsequent levels. We performed this contrast on part of the data of Experiment 1 in Maes et al. (2004), using the post-shift mean number of errors from Conditions P, LI, and P+LI, and taking Condition P as the first (the basis for the comparisons because it did not differ from the control conditions) and Condition P+LI as the last condition, respectively. The Helmert contrast revealed a highly significant effect for the first comparison (Condition P vs. mean of Conditions LI and P+LI: $p = .002$, partial $\eta^2 = .25$), but not for the second comparison (Condition LI vs. Condition P+LI: $p = .27$, partial $\eta^2 = .04$). The question of interest was whether an identical contrast would yield similar results for the present data set.

3. Results

The mean number of pre-shift trials, including the final eight correct trials, was 32.3 for Condition P, 29.8 for Condition LI, and 38.8 for Condition P+LI. The corresponding number of post-shift trials were 23.8, 39.9, and 44.7.

The conditions did not differ much in the mean number of errors in the pre-shift phase, 12.5, 9.3, and 12.0, for Con-

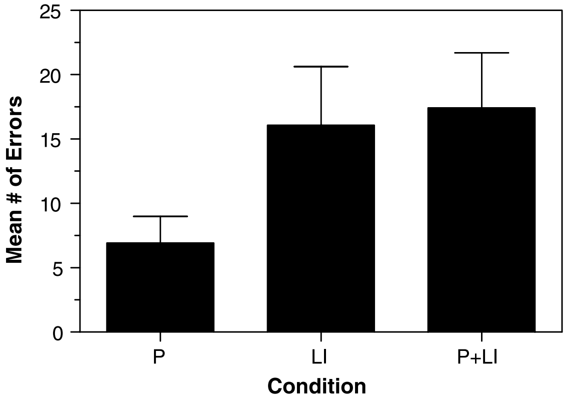

Fig. 2. Mean (+SEM) number of post-shift errors for each experimental condition. Participants in Conditions P and LI could only make post-shift errors based on, respectively, response perseveration and learned irrelevance. Participants in Condition P+LI could make both types of error.

ditions P, LI, and P+LI, respectively. These latter differences clearly were not significant, ANOVA with Condition as single factor, $F < 1$.

Fig. 2 displays the mean number of post-shift errors for each condition.

Although ANOVA with Condition as factor on the data shown in Fig. 2 failed to reveal a significant effect, $F(2,33) = 2.27$, $p = .12$, also due to the relatively large variance in Conditions LI and P+LI, the more sensitive planned Helmert contrast using the post-shift data did reveal a significant difference between Condition P and the mean of Conditions LI and P+LI, $p < .05$, partial $\eta^2 = .12$, but not between Conditions LI and P+LI, $p > .8$, partial $\eta^2 = .00$.

4. Discussion

This study is the first to explicitly test the relative contribution of errors due to learned irrelevance and to response perseveration in healthy participants subjected to a total change dimensional-shift task. The results suggest a larger role for learned irrelevance than for response perseveration. A similar finding was obtained in a previous study that differed from the present one in one important aspect: the exemplars of primary interest used in the pre- and post-shift phases were not changed in that study (Maes et al., 2004). Therefore, the present results provide evidence for the generality of the observed predominance of learned irrelevance.

An important benefit of a total change procedure is that it eliminates the possibility that differences between conditions reflect differences in, for example, the opportunity to detect the shift, or in the obviousness of post-shift solutions (see Slamecka, 1968, for a detailed discussion). For example, it may be argued that in Condition P of the Maes et al. (2004) study, the subjects had to test a smaller number of exemplars in the post-shift phase than was the case in Condition LI. After the first negative feedback in the post-shift phase in Condition P, the subjects could be prompted to first test the former two incorrect exemplars of the previous relevant dimension, followed by a test of all four exemplars of the novel dimension. For the subjects in Condition LI, it could be argued that the very first negative feedback in the post-shift phase would first let them to evaluate all four exemplars of the novel (but irrelevant) dimension, followed by a test of all four exemplars of the previously irrelevant (and thus ignored) dimension. Therefore, the former condition implies fewer alternatives to test than does the latter condition. Such a potential confound is absent in a total change procedure because of the novelty of all exemplars in each condition.

Perhaps, the most parsimonious interpretation of the common findings of our present and previous study is that humans always approach dimensional shift tasks in terms of abstract rules and dimensions, regardless of whether or not the same exemplars are used. Accordingly, it is ‘dimensions’ that are subject to learned irrelevance and perseveration mechanisms. A somewhat weaker version of this position is that, in each type of procedure, subjects first learn to direct their attention to the relevant dimension, which may be considered to reflect a higher-order cognitive process, and subsequently learn about exemplar-feedback associations. This approach is far from novel (e.g., Sutherland & Mackintosh, 1971), and can account for a number of other discrimination learning phenomena, such as the faster learning of intra-dimensional shifts as opposed to extra-dimensional shifts frequently observed in both humans and non-humans. It could, in principle, also account for the fact that the difference between conditions was somewhat larger in the Maes et al. experiments than in the present experiment (e.g., compare the effect size for the contrast between Condition P and the mean of Conditions LI and P+LI in the present study, partial $\eta^2 = .12$, and in the previous study, partial $\eta^2 = .25$) by assuming that learned irrelevance can take place at the level of both dimensions and exemplars in the previous experiments, but only at the level of dimensions in the present experiment.

Another theoretical possibility is that both types of task exclusively evoke associative learning processes at the exemplar level. To explain transfer of learning about exemplars in the pre-shift phase to learning about novel post-shift exemplars, one could adopt the notion of generalization. Accordingly, it may be argued that increased or decreased attention to, and associative strength of, one stimulus generalize more to other exemplars of the same dimension than to exemplars from another dimension (e.g., see Mackintosh, 1975). In essence, this account would imply identical learned irrelevance and perseveration mechanisms in both types of task (namely at the exemplar level), as, for example, formally described in some associative learning models (Mackintosh, 1975; Pearce & Hall, 1980). The difference between tasks would merely be quantitative, with these processes being attenuated in the present task (due to generalization decrement) compared to the previous task. Perhaps, the ‘exemplar versus dimension’ issue is just a matter of level of description. What is described at a ‘cognitive’ level of description in terms of a dimension is ultimately represented at the neuronal level in terms of, for example, exemplars generating activation patterns that are distributed across a set of nodes, with different exemplars activating partially overlapping sets, and exemplars belonging to the same dimension having a larger overlap than exemplars from different dimensions (e.g., see McLaren & Mackintosh, 2002).

To conclude, the important empirical finding of the present study is a larger influence of learned irrelevance than perseveration in healthy subjects, even if these processes could not directly operate on specific exemplars. This predominance, therefore, seems to be a rather general finding that should be taken into account when evaluating task performance of clinical populations. For example, the observation that a specific clinical population primarily displays errors due to perseverative responding rather than to learned irrelevance is even more noteworthy given the predominance of the learned irrelevance mechanism in healthy subjects.

References

Barceló, F., & Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia, 40*, 349–356.

Gauntlett-Gilbert, J., Roberts, R. C., & Brown, V. J. (1999). Mechanisms underlying attentional set-shifting in Parkinson's disease. *Neuropsychologia, 37*, 605–616.

Joosten, J. P. A., Coenders, C. J. H., & Eling, P. A. T. M. (1995). Shifting behavior: an analysis of response patterns of Parkinson patients in discrimination learning. *Brain and Cognition, 29*, 115–126.

Lanser, M. G., Berger, H. J. C., Ellenbroek, B. A., Cools, A. R., & Zitman, F. G. (2002). Perseveration in schizophrenia: failure to generate a plan and relationship with the psychomotor poverty subsyndrome. *Psychiatry Research, 112*, 13–26.

Lawrence, A. D., Sahakian, B. J., Rogers, R. D., Hodges, J. R., & Robbins, T. W. (1999). Discrimination, reversal, and shift learning in Huntington's disease: mechanisms of impaired response selection. *Neuropsychologia, 37*, 1359–1374.

Li, C.-S. R. (2004). Do schizophrenia patients make more perseverative than non-perseverative errors on the Wisconsin Card Sorting Test? A meta-analytic study. *Psychiatry Research, 129*, 170–190.

Mackintosh, N. J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276–298.

Maes, J. H. R., Damen, M. D. C., & Eling, P. A. T. M. (2004). More learned irrelevance than perseveration errors in rule shifting in healthy subjects. *Brain and Cognition, 54*, 201–211.

McLaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. *Generalization and discrimination. Animal Learning and Behavior, 30*, 177–200.

Owen, A. M., Roberts, A. C., Hodges, J. R., Summers, B. A., Polkey, C. E., & Robbins, T. W. (1993). Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or Parkinson's disease. *Brain, 116*, 1159–1175.

Pantelis, C., Harvey, C. A., Plant, G., Fossey, E., Maruff, P., Stuart, G. W., Brewer, W. J., Nelson, H. E., Robbins, T. W., & Barnes, T. R. E. (2004). Relationship of behavioural and symptomatic syndromes in schizophrenia to spatial working memory and attentional set-shifting ability. *Psychological Medicine, 34*, 693–703.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*, 532–552.

Ravizza, S. M., & Ciranni, M. A. (2002). Contributions of the prefrontal cortex and basal ganglia to set shifting. *Journal of Cognitive Neuroscience, 14*, 472–483.

Slamecka, N. J. (1968). A methodological analysis of shift paradigms in human discrimination learning. *Psychological Bulletin, 69*, 423–438.

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.