Chapter 8
Learned Irrelevance Revisited: Pathology-Based Individual Differences, Normal Variation and Neural Correlates

Aleksandra Gruszka, Adam Hampshire, and Adrian M. Owen

Introduction

The function of the human executive system can broadly be described as the seeking out and processing of those signals and memories that are of the greatest relevance when guiding deliberate and adaptive behaviours. This task is not easy, however, since it requires almost constant shifting of attention in response to irregular alterations in the contingencies relating stimuli, responses, and environmental feedback. An individual’s current belief regarding these contingencies guides response within a given context, and the representation of this belief and its consequent behaviour is often referred to as an “attentional set”. Consequently, attentional set-shifting is an important executive function responsible for altering a behavioural response in reaction to the changing contingencies (Cools, Barker, Sahakian, & Robbins, 2001; Gotham, Brown, & Marsden, 1986). Such flexibility underlies a wide range of behaviours: the better the set-shifting capacity, the more flexible the person is at adapting to change. At the other end of this continuum are many psychiatric groups, neurodegenerative groups and even healthy elderly and young subjects that have been shown repeatedly to be impaired in attentional set-shifting performance. One specific form of these impairments lies in an inability to attend to, or to learn about, information which has previously been shown to be irrelevant. This phenomenon called learned irrelevance (LI) (Mackintosh, 1973) is very mysterious, because unlike other aspects of attentional set-shifting, it appears to be neither dependent on the frontal lobe (e.g. Owen et al., 1993) nor affected by dopamine (Owen et al., 1993; Slabosz et al., 2006), and, therefore, may not be coded for in the parts of the brain that are typically considered “executive” at all.

The aim of this chapter is to discuss recent advances in the area of LI in humans and to show how the latest trends in executive-function research can be applied to the study of LI. The first trend is the application of experimental paradigms that provide measures of putative cognitive functions much more precisely than the measures that have been offered by “classical” neuropsychological methods (Aron, 2008). One such paradigm is the ID/ED visual discrimination learning paradigm (e.g. Downes et al., 1989; Owen et al., 1992) modelled after the most prominent neuropsychological tool for studying attentional set-shifting deficits, namely Wisconsin Card Sorting Task (WCST; Grant & Berg, 1948). The ID/ED paradigm allows the operationalization of the dependent variables in a much more reliable way. It makes them much more sensitive to the effects of brain damage or...
pathophysiology. In the first section of this chapter, we review studies that have utilised the ID/ED paradigm to investigate normal and pathology-based individual differences in LI. The utility of such analyses is motivated by the fact that dissociable patterns of LI deficits were observed in patients with circumscribed frontal-lobe removals, and both medicated and non-medicated patients with Parkinson’s disease (PD). An understanding of these patterns may lead us to unravel the distinct roles played by different frontal striatal circuits (Alexander, Delong, & Strick, 1986), or the different roles played by the cortical and striatal portions of those circuits.

The second trend in executive-function research is the implementation of advanced neuroscience techniques, such as brain imaging (Aron, 2008). In the second section of the present chapter, we describe our study that aimed to investigate the neural mechanisms underlying dissociable components of attentional set-shifting, including LI. Taken together, the combined approach based on the two trends enables a finer delineation of executive functions than was possible with classical paradigms. In our opinion, this combined approach will ultimately allow us to answer the question of how LI effect is rooted in actual neural systems. One may say that this chapter exemplifies a situation where individual differences are less important as such; rather, the dissociable patterns of results obtained for various neuropsychological groups and the normal population are more significant, since these dissociable patterns may help us understand the neural and neurochemical mechanisms underlying the investigated phenomena.

Learned Irrelevance and a Visual Learning Paradigm

Studies of human executive functions first began when several tests of behavioural flexibility, now considered classic, were administered to patients with frontal lobe lesions. Among these tests, the WCST was given a particularly prominent place in standard neuropsychological testing of attentional set-shifting ability. The WCST requires participants to sort a deck of cards according to a number of dimensions. Each card varies in three dimensions – number, colour and shape – and as each card is presented, the subject is required to match it according to a specified dimension. After a certain number of correct “sorts”, the rule is changed and the subject is required to start sorting according to an alternative dimension. In order to comply with the new task requirements, subjects have to override a tendency to stick to a previously relevant rule. It has been shown repeatedly that damage to various regions of the frontal lobe produces behavioural impairment on the WCST, particularly for conditions in which the subject had to overcome a prepotent response tendency. However, the results indicating that impaired performance on the WCST is linked to frontal-lobe dysfunction have often been inconsistent. In separate studies, dorsolateral (Demakis, 2003), orbitofrontal (Dias, Robbins, & Roberts, 1997) and medial (Drewes, 1974) brain areas have been implicated in various aspects of WCST performance. In fact, it has been argued that there is no clear support for the role of the WCST as a diagnostic tool of frontal lobe damage (Mountain & Snow, 1993; Reitan & Wolfson, 1994), since diffuse brain injury (Fork et al., 2005) as well as localised damage to specific non-frontal regions (Anderson, Bigler, & Blatter, 1995; Canavan et al., 1989; Hermann, Wyler, & Richey, 1988; Horner, Flashman, Freides, Epstein, & Bakay, 1996) can produce similar cognitive impairments. In addition, several cases have been reported in which there was a lack of an impairment in spite of frontal-lobe pathology (Anderson, Damasio, Jones, & Tranel, 1991; Eslinger & Damasio, 1985).

Some of the inconsistency is probably related to the low psychological resolution of the WCST. Completing the test involves the recruitment of a number of different executive functions (e.g. performance monitoring, integration of feedback, rule-induction, set-shifting, and suppression of previous sorting rules), and these may be coded in discrete circuits – some of which may not be within the frontal lobes. The WCST requires subjects to shift attentional set away from the previously relevant dimension. The term “set”, used in this context, refers to a predisposition to attend selectively to a particular stimulus dimension (such as “colour” or “shape”), established on the basis of reinforcing feedback (i.e. “correct”
incorrect" cues). However, there are many forms of shifting implicit in the WCST. For example, a new rule may require that subjects shift attentional-set either to the other so-far "incorrect" exemplar of the line dimension (e.g. from red to blue) or to shift their attention to the other so-far "incorrect" dimension (e.g. from shape to colour). Studies with humans and experimental primates have suggested that these two forms of shifting — extra-dimensional (ED) shifting and reversal shifting, respectively — are subserved by dissociable regions within the frontal cortex. While EDS shifting is hampered by damage to the dorsolateral prefrontal cortex (DLPFC), reversal shifting is impaired by lesions to the OFC or ventromedial frontal cortex (and not the DLPFC) (Dias, Robbins, & Roberts, 1996; Fellows & Farah, 2003; Jones & Mishkin, 1972). As a result, using the WCST it has not been possible to define a specific area within the frontal cortex that is critically involved in attentional set-shifting and several recent studies (Downes et al., 1989; Owen et al., 1992; Roberts, Robbins, & Everitt, 1988) have argued that pure “set-shifting ability” may be more accurately assessed using the intra- and extra-dimensional shifting paradigm described frequently in animal learning literature (Mackintosh, 1983). An “intradimensional shift” (IDS) occurs when a subject is required to cease responding to one exemplar of a particular stimulus dimension (e.g. “blue” from the dimension “colour”) and must begin responding to a new exemplar of that same dimension (e.g. “red”). As mentioned above, an “extradimensional shift” (EDS) occurs when a subject is required to switch responding to a novel exemplar of a previously irrelevant dimension (e.g. from “blue” to “squares” from the dimension “shape”). In discrimination learning tasks, impairments in neuropsychological or neurological populations are observed mainly when an EDS shift is required, rather than when an IDS shift is required (Downes et al., 1989; Roberts et al., 1988). In one study, a group of neurosurgical patients with frontal-lobe lesions were shown to be specifically impaired in their ability to shift response set to the previously irrelevant stimulus dimension (i.e. at the EDS stage of learning) but not to shift attention to new exemplars of a previously relevant dimension (i.e. at the IDS stage of learning) (Owen, Roberts, Polkey, Sahakian, & Robbins, 1991). By comparison, patients with medial temporal-lobe excisions were not impaired in their ability to perform either shift.

Furthermore, EDS deficits in attentional set-shifting have been dissected further into two separate contributing mechanisms. It has been suggested that these deficits may reflect either an impairment in the ability to shift attention from a perceptual dimension that has been previously relevant (i.e. perseveration”), or in the ability to shift to an alternative perceptual dimension that has been previously perceived as irrelevant (i.e. “learned irrelevance”) (Owen et al., 1993). This distinction plays an important role in the context of neuropathology (Owen et al., 1993), since separate cognitive and possibly neurochemical mechanisms of perseveration and LI have been suggested (Maes, Dannen, & Eling, 2004; Owen et al., 1993).

The phenomenon of perseveration seems much better understood, both in terms of the underlying cognitive operations and the neuroanatomical structures, despite the fact that existing studies imply a predominance of the LI over perseveration mechanism in the normal population (Maes, Vich, & Eling, 2006; Maes et al., 2004). Perseverative behaviour on WCST-like tasks may result from several types of endogenous adaptive control errors, i.e. failures in rule induction (i.e. forming hypotheses concerning the new rule), inability to shift attentional set (the configuration of appropriate task sets to test these hypotheses and the suppression of no-longer-relevant task sets), or deficient monitoring of performance (Riddervold, Span, & van der Molen, 2002). As for the neural correlates of perseverative behaviour, it appears to be a general consensus in clinical studies, experimental neuropsychology and cognitive neuroimaging that perseverative behaviour mainly reflects inefficient prefrontal function (see Barcelo, Sanz, Molina, & Rubio, 1997 for a review). Brenda Milner (1963) was the first author to link poor WCST performance with circumscribed DLPFC lesions rather than OFC or more posterior lesions. In addition, she concluded that patients with frontal-lobe lesions were more susceptible to “perseverative errors”, i.e. persisting responding according to the previously relevant rule despite continued negative feedback. Since Milner’s report (1963) the specific relationship between perseverative behaviour and deficiencies in frontal cortex functioning has been confirmed many times (e.g. Barcelo et al., 1997; Barcelo & Santome-Calleja, 2000; Drewe, 1974; Stuss & Benson, 1984). Recent neuroimaging studies also suggest
that overcoming perseverative tendencies correlates with activity in prefrontal regions (Nagahama, Okina, Suzuki, Nabatame, & Matsuda, 2005).

In contrast, the paucity of reports on LI in humans makes it poorly understood, both in terms of cognitive processes and neural/neurochemical mechanisms that underlie it. In cognitive terms, the phenomenon is clearly related to latent inhibition. However, whereas LI refers to impaired learning of an association between a conditioned stimulus and an unconditioned stimulus because of their uncorrelated presentations (Mackintosh, 1983), latent inhibition refers to disrupted learning following unreinforced presentations of the conditioned stimulus alone (Lubow, 1973; see Lubow, this volume). Several theories of latent inhibition can be used to account for LI (Gluck & Myers, 1993; Lubow, 1989; Weiner & Feldon, 1997). According to some authors, LI may be a special case of latent inhibition, occurring as a result of pre-exposure of the unconditioned stimulus and conditioned stimulus (Bonardi & Hall, 1996). According to others, however, LI is inexplicable as a simple summation of the two pre-exposure effects and cannot always be reduced to latent inhibition (Baker & Mackintosh, 1979; Bennett, Wills, Oakeshott, & Mackintosh, 2000; Matzel, Schachtman, & Miller, 1988). Instead, it is argued that explicit learning occurs about the absence of a correlation between the conditioned stimulus and the unconditioned stimulus. The outcome of this process then interferes with further learning about a subsequent positive correlation.

Because the relationship between LI and latent inhibition is so unclear, we decided to limit our analyses to those cases of LI described in the literature, that come from research based on a visual learning paradigm. In other words, in this way we operationalize our understanding of LI, to avoid comparing these two concepts and paradigms any further. Hoping that this approach may offer new understanding of individual differences in executive functioning, we will now review the evidence for individual differences in LI in various clinical populations, with particular emphasis on PD as the most extensively studied neuropsychological population in this context, and individual differences in LI as a function of age.

Pathology-Based Individual Differences in Learned Irrelevance

Parkinson’s Disease

Idiopathic Parkinson’s disease (PD) is a neurodegenerative disorder, which is clinically defined on the basis of a triad of motor symptoms: bradykinesia, rigidity and resting tremor. Nevertheless, cognitive impairments are a common characteristic of the condition and strong predictors of quality of life in such patients. Some of these impairments closely resemble frontal lobe deficits, including problems in shifting attentional set in both cognitive and motor domains (Cools, van den Bercken, Horstink, van Spaendonck, & Berger, 1984; Downes et al., 1989; Owen et al., 1992; van Spaendonck, Berger, Horstink, Buytenhuijs, & Cools, 1996). Numerous studies from the cognitive domain have shown that attentional set-shifting is impaired in the early stages of PD (e.g. Downes et al., 1989; Owen et al., 1992) and furthermore, it has been suggested that attentional set shifting is moderately selective to PD (e.g. deficits are not seen in early Alzheimer’s disease; see Sahakian et al., 1990).

Attentional set-shifting performance in the patients with PD has been studied most extensively with the tests of visual discrimination learning described in the previous section (e.g. Downes et al., 1989; Owen et al., 1992). Using such tests, a number of studies have shown that PD patients are more impaired when a shift of attention is required between two different perceptual dimensions, i.e. EDS, than when a shift is required between two different values of the same dimension, i.e. IDS (Downes et al., 1989). However, no consensus has been reached regarding the mechanisms underlying the poor performance of PD patients. To account for these deficits, various cognitive mechanisms have been suggested, including impairments in set-shifting (e.g. Taylor, Saint-Cyr, & Lang, 1986).
maintaining set (Flowers & Robertson, 1985; Taylor et al., 1986) or concept formation (Cooper, Nagar, Jordan, Harvey, & Sullivan, 1991).

As for LI, both medicated and non-medicated patients with PD (Downes et al., 1989) with mild or severe motor symptoms related to different stages of the condition (Owen et al., 1992) have been shown to be impaired on a test requiring shifting of attentional set. In a further study by Owen et al. (1993), the performance of medicated and non-medicated patients with PD was compared to that of patients with frontal-lobe lesions whilst undertaking a test that independently manipulated the perseveration and LI aspects of ED shifting. In the perseveration condition, patients were presented at the EDS with one dimension that was already familiar and had always been relevant to the task rule during previous stages of learning and with a second dimension that was novel. They were required to shift response set to the novel dimension, ignoring the previously relevant dimension (i.e. attempt to overcome the tendency to perseverate). In the LI condition, patients were presented at the EDS with one dimension that was already familiar but had been irrelevant during all previous stages of learning and with a second dimension that was novel. They were required to shift response set to the familiar (but previously irrelevant) dimension, ignoring the novel dimension (i.e. attempting to overcome LI).

In the study by Owen et al. (1993), the patients with frontal lobe excisions committed the highest number of errors in the perseveration condition out of the groups participating in the study (i.e. the control group and the two PD groups – medicated and non-medicated). At the same time, the frontal-lobe group performance on LI was comparable to the performance of the healthy control group in this condition. In contrast, non-medicated patients with PD exhibited problems in both the perseveration condition and the LI condition, while medicated patients with PD were impaired only at the LI condition. These results have several major implications for understanding the nature of attentional set-shifting deficits both in PD and as an executive function in general. First, they suggest that the major set-shifting deficits reported in both patients with PD and frontal lobe patients may involve fundamentally different, though related cognitive processes, with both perseveration and LI contributing to the cognitive impairments observed in PD, and perseveration, but not LI, contributing to set-shifting impairments in frontal-lobe patients. Second, perseveration, but not LI, responds to L-dopa therapy, suggesting that the former, but not the latter, is related to the central dopaminergic deficit in PD.

However, various aspects of the LI hypothesis as a mechanism accounting for attentional set-shifting deficits in PD remained controversial. Van Spaendonck et al. (1995) were not able to reproduce reliable LI impairments in patients with PD with a classical card sorting design, and hence, they have attributed attentional set-shifting deficits in PD to problems with self-generation of problem solving strategies. Similarly, Gauntlett-Gilbert, Roberts, and Brown (1999) failed to show that the performance of the patients with PD (vs. controls) was preferentially improved under the circumstances where strong LI acquired during a pre-shifting stage of a visual discrimination learning task should actually facilitate the performance at the subsequent shifting phase.

Recently, in an attempt to better define the cognitive and neural basis of the learnt irrelevance phenomena, we have developed a novel visual discrimination learning test, which allowed for a finer delineation of factors responsible for LI by avoiding the introduction of a novel dimension at the EDS stage of the test (Slabosz et al., 2006). This manipulation prevented the possibly confounding effect of novelty on the LI measure.

Instead of introducing a novel dimension, the authors varied the extent to which a target dimension was irrelevant prior to the EDS, and the participants were required to shift attentional set to a dimension that had been either fully irrelevant or partly reinforced (see Fig. 8.1). Prior to the EDS stage of the test, the stimuli were characterised by three dimensions: colour, shape and number of items; and up to the EDS stage only one of the three dimensions (colour) was relevant to the discrimination rule and consistently reinforced. At the same time, the level of task irrelevance of the other two dimensions was varied, with one dimension (either shape or number) being fully irrelevant, and the other dimension being partly reinforced. In the case of the fully irrelevant dimension, any given value of this dimension (e.g. square or circle) randomly co-occurred with the reinforced
Fig. 8.1. Learned irrelevance: summary of the procedure for the IDS and EDS stages of the learned irrelevance task. Stimuli shown are only examples (from Slabosz et al., 2006).

value of the currently relevant dimension (i.e. blue or red). In other words, the fully irrelevant dimension was reinforced randomly and in this sense was equivalent to the irrelevant dimension of the original CANTAB ID/ED task. In contrast, in the case of the partially relevant dimension, one exemplar co-occurred with the reinforced value of the currently relevant dimension on 75% of trials preceding the EDS. As a result, the partially relevant dimension predicted the reinforcement at a level that was greater than chance. At the EDS stage of the task, the participants were required to shift their attention either to the previously irrelevant dimension (the full irrelevance condition) or to the dimension that had previously been partially reinforced (the partial relevance condition). The results revealed that patients with PD made more errors than control participants in the fully irrelevant condition but not in the partially relevant condition (see Fig. 8.2). Moreover, L-dopa had no effect on the patients’ task performance, despite improving their working memory - as shown in a separate control task. These results confirm that LI is a significant factor accounting for attentional set-shifting deficits in patients with PD, although unlike other executive impairments in this group, the phenomenon appears to be unrelated to their central dopaminergic deficit.

These findings (Slabosz et al., 2006) have advanced our understanding of LI in two important ways. First, in previous studies of LI (Gauntlett-Gilbert et al., 1999; Owen et al., 1993), a significant confound has been the possible effects of stimulus novelty on EDS performance. That is to say,
shifts to a previously irrelevant dimension were compared with shifts to a novel dimension, to make inferences about LI. This is important because deficits in novelty pop-out effects have been reported previously in patients with PD (Lubow, Dressler, & Kaplan, 1999; Tsuchiya, Yamaguchi, & Kobayashi, 2000). In our recent study (Slabosz et al., 2006), no such confound existed, as inferences about LI were based on comparisons between full and partial irrelevance conditions, neither of which involved the introduction of a novel dimension. Second and more importantly, the research by Slabosz et al. (2006) and related studies have suggested that LI appears to be neither dependent on the frontal lobe (e.g. Owen et al., 1993) nor affected by dopamine and therefore, may not be executive at all. Thus, the question that remains is whether a plausible neural and/or neurochemical account can be formulated for the phenomenon of LI.

**Pathologies Other Than PD**

Unfortunately, relevant data from other clinical groups is sparse, although the fact that patients with circumscribed excisions of the frontal cortex are unaffected on LI tasks implicates mechanisms other than those that are traditionally considered to be executive. Although patients with schizophrenia have recently been reported as showing abnormal performance during a test of LI, the pattern of impairments suggests a reduced rather than an enhanced effect (Gal et al., 2005). Thus, among first-episode schizophrenic patients, cue-target associations to (irrelevant) pre-exposed cues were as fast as those to novel cues (see also Gal et al., 2005), exactly the opposite pattern to that which would be predicted in PD on the basis of the previous findings (Owen et al., 1993; Slabosz et al., 2006). Although such evidence may suggest a role for dopamine in LI, the direct manipulation of dopamine levels through medication conducted in the study by Slabosz et al. (2006) more strongly suggests otherwise. In fact, the lack of effect of L-dopa and of frontal-lobe damage (Owen et al., 1993) on LI in patients with PD suggests that neither the dopaminergic mechanisms of the striatum nor the prefrontal cortex mediate this process. In monkeys, prefrontal dopamine depletion impairs spatial working memory but has no significant effect on extradimensional set-shifting performance (Roberts et al., 1994), a result that is broadly consistent with those of the study by Slabosz et al. (2006). Although
pathologies affecting neurotransmitters other than dopamine in PD have been proposed, including noradrenergic, serotonergic and cholinergic deafferentation of the cortex (Agid, Javoy-Agid, & Ruberg, 1987), the dopamine hypothesis remains a leading factor in understanding PD.

Normal Variation in Learned Irrelevance

Age-Related Variation in Learned Irrelevance

Unfortunately, not much is known about the modulatory role of age on susceptibility to LI, since to our knowledge only one study has examined this issue directly. Slabosz et al. (2006) compared the performance of elderly and young volunteers on the LI task described above. The mean ages of the two groups were 35.9 years and 68.5 years, respectively. This comparison revealed no significant differences for the cross-group comparisons. Thus, aging did not seem to influence LI performance. However, this conclusion needs to be treated with caution and has yet to be confirmed. Moreover, this observation contrasts markedly with the well-recognised result that aging enhances susceptibility to perseveration (Foldi, Helm-Estabrooks, Redfield, & Nickel, 2003). Bearing in mind that frontal lobe deficiencies are considered to be the main age-related neural pathology (Hampshire, Gruszka, Fallon, & Owen, 2008), this result suggests again that the mechanisms of LI may not be frontal at all.

To sum up, a differential pattern of susceptibility to LI has been observed in the frontal-lobe patients, patients with PD who are “on” dopaminergic medication, and PD patients who are “off” medication. While the frontal-lobe patients seem to be able to overcome LI, both groups of patients with PD are comparably susceptible to it. Similarly, aging does not seem to affect susceptibility to LI. This all suggests that mechanisms other than those that are dependent on the frontal cortex and those that are related to the dopaminergic system are responsible for LI. This coherent picture is somehow blurred by the recent findings revealing that patients with schizophrenia exhibit a reduced LI effect (Young et al., 2005). Although this pattern of results would suggest a dopaminergic basis for LI (given the known dopaminergic pathology in schizophrenia), the study by Slabosz et al. (2006) designed specifically to test this hypothesis strongly suggests otherwise. Thus, on the basis of clinical data alone, it has not been possible to define specific neural or neurochemical mechanisms that are critically involved in LI.

Neural Correlates of Learned Irrelevance

A variety of neuroimaging and neuropsychological studies have led to the view that the neural circuitry responsible for co-ordinating attentional set-shifting consists of many parts, including the dorso-lateral and ventrolateral prefrontal cortices, the anterior cingulate, and the posterior parietal cortex. Whilst the existence of this network is no longer a matter of debate, the contributions made by these anatomically distinct components are, as yet, poorly defined. Much of the current confusion regarding the precise nature of frontoparietal organisation results from the use of complex and cognitively heterogeneous task manipulations when attempting to functionally dissociate frontoparietal influences (see Hampshire & Owen, 2006, for discussion) – reflecting the confusion in the behavioural research on attentional set-shifting as discussed above.

Studying LI poses even more problems for neuroimaging because LI is usually demonstrated as a one-off, between-groups difference in trials involving learning to perform an EDS. In other words, the LI paradigms that to date have been used to examine behavioural effects are unsuitable for the repeated measurements that are a prerequisite for fMRI analysis (Owen et al., 1993; Slabosz et al., 2006) due to the fact that the dependent measure is the percentage of subjects who passed to the next (e.g. EDS) stage.
of the task (e.g. Owen et al., 1992) or completed the number of trials needed to learn the EDS (e.g. Slobos et al., 2006). An additional limitation to these conventional LI paradigms is that they use a between-subject design (e.g. Owen et al., 1993) and, as a result, an LI effect can be determined only by group comparisons. Thus, to our knowledge, only one study has investigated neural correlates of LI directly (Young et al., 2005), having demonstrated involvement in LI of various components of the hippocampal formation, in agreement with studies on latent inhibition. Studies with experimental animals have also shown a key role played by limbic structures in mediating latent inhibition (Gray et al., 1995, 1997). Dopaminergic function in the mesolimbic pathway, terminating in the nucleus accumbens, appears to be critical (Joseph et al., 2000), as does activity in the hippocampal formation and entorhinal cortex (Honey & Good, 1993). However, Young et al. (2005) have used a different paradigm (i.e. cue-target association learning) to the visual learning paradigm discussed here and their findings derive from a relatively small sample, which needs replication.

The above mentioned limitations of imaging studies of shifting attentional set were addressed by a recent fMRI study by Hampshire and Owen (2006). In this study, a novel approach was used: the responses of the volunteer dictated the order and pace of experimental events. In this way, it was not the experimental design but the focus of attention that defined the events that were used in the fMRI model (e.g. attentional shifts). The chosen decision-making strategies and attentional shifts were thus functionally and behaviourally examined independently of the will of the experimenter. The study used many stimulus sets each containing stimuli of two distinct types (faces and buildings) and modelled switches of attention between stimuli of the same type (IDS) and between stimuli of different types (EDS). Hampshire and Owen (2006) intermixed these transient attentional control functions that could therefore be contrasted at the event level in the current trial and error situation. Accordingly, EDS and IDS were compared directly, effectively isolating the extra-dimensional component of shifting from other switch-related processes, such as inhibition of the previously relevant response (Nakahara, Hayashi, Konishi, & Miyashita, 2002). Switch and feedback events were modelled separately thanks to the use of the novel partial feedback paradigm that allowed regions involved in abstract reward processing and/or the implementation of attentional control to be modelled separately.

The behavioural data showed that moving attention between stimulus dimensions caused more errors, and took longer, than moving attention between stimuli of the same type. This difference must reflect the typical strategy employed by the volunteers to solve the task since both extra- and intra-dimensional target changes could logically be solved within the same number of trials. The main query for the imaging data, therefore, was whether this component of attentional control (EDS) would be associated with any specific neural substrate. Accordingly, when shifts in the focus of attention between stimulus types (EDS) were directly compared with shifts within stimulus type (IDS), significant activation was seen only in the VLPFC and the preSMA. Based on these findings, the authors suggested that the commonly observed increase in reaction time for extradimensional shifting reflects the time taken for the ventrolateral frontal cortex to bias attentional processing between competing stimulus dimensions (Hampshire & Owen, 2006). Such attentional biasing, or “tuning”, while relevant to many components of set-shifting, is probably maximal when a complete reconfiguration of the attentional set is required, as is the case during a shift from one dimension to another (competing) dimension.

Recently, the paradigm developed by Hampshire and Owen (2006) has been adapted to include LI and perseveration aspects of EDS performance (Gruszka, Hampshire and Owen, in prep.). The procedure of the current task differed from the original design (Hampshire & Owen, 2006) in several important ways in order to resemble the LI/perseveration paradigm proposed by Owen et al. (1993) more closely. First of all, there were several abstract dimensions introduced in a course of the task (i.e. shape, colour, number, spatial location of the pattern, texture, instead of faces and buildings) and subjects were required to switch between these dimensions in a pseudo-random manner. Hence, when the stimulus set was changed after a criterion number of continuous correct selections was achieved, one novel dimension was introduced (i.e. absent before this change), and
Fig. 8.3 Learned irrelevance and perseveration: illustration of a typical series of trials. Stimuli shown are for example only (from Gruszka et al., in prep.)
one previously present dimension remained (either previously relevant or previously irrelevant dimension; see Fig. 8.3). This resulted in four main task conditions that were defined by the dimension that the selected exemplar belonged to when a new set of stimuli was presented. These were: (1) LI succumbing condition (i.e. choosing the novel dimension when the previously irrelevant dimension and the novel dimension were present), (2) LI overcoming condition (i.e. choosing the previously irrelevant dimension when the previously irrelevant dimension and the novel dimension were present), (3) PE succumbing condition (i.e. choosing the previously relevant dimension when the previously relevant dimension and the novel dimension were present), and finally (4) PE overcoming condition (i.e. choosing the novel dimension when the previously relevant dimension and the novel dimension were present). This allowed events related to LI and PE to be contrasted directly. The behavioural data showed that in agreement with expectations, shifts of attention under both LI and PE “succumbing” conditions were easier for the subject than those requiring overcoming LI or PE, respectively (Fig. 8.4).

Overall, the imaging results of the pilot study reported here confirmed the results obtained by Hampshire and Owen (2006) with a broad activation of the fronto-parietal network when comparing working out versus knowing the target (see Fig. 8.5). This suggests that in general, the paradigm is suitable for studying more complex attentional set-shifting activities than the simple two-dimensional task used by Hampshire and Owen (2006). However, the results of the EDS-IDS contrast differ substantially from those obtained by Hampshire and Owen (see Fig. 8.6), with a greater proportion of the frontoparietal network recruited when switching attention between stimulus dimensions. Note that this activation included a significant DLPFC component. Previously, in a block-design positron emission tomography (PET) study, Rogers, Andrews, Grasby, Brooks, and Robbins (2000) have shown increased activity in DLPFC, but not VLPFC, during EDS when compared to IDS. However, unlike the study by Hampshire and Owen (2006), it is likely that in both Rogers et al. (2000) and the current study, the activation observed could well have been due to the additional

Fig. 8.4 Learned irrelevance and perseveration: the effects on the number of errors made while searching for the target under learned irrelevance and perseveration overcoming/succumbing conditions (from Guszka et al., in prep.)
demands of actively working out which dimensions were relevant to the task, rather than the more specific shift of attention between dimensions. This result highlights a possible dissociation between the VLPFC and the DLPFC, with the former involved in routine switches of attention within a well defined attentional set and the latter involved in identifying which rules and dimensions are most relevant for forming that attentional set.

As to LI and PE components of the task, there were two main contrasts of interests. The first was the comparison of the two LI conditions: overcoming vs. succumbing at the moment of set change (i.e. choosing the previously irrelevant dimension when the previously irrelevant dimension and the novel were present minus choosing the novel dimension when the previously irrelevant dimension and the novel dimension were present). The second contrast was the comparison of both LI conditions (overcoming and succumbing) minus both PE conditions (overcoming and succumbing) at the moment of set change (i.e. taking the new set when the previously irrelevant dimension and the novel were present minus when the previously relevant dimension and the
novel dimension were present). At this preliminary stage, the data were examined with both the exploratory ROIs analyses using a set of anatomical ROIs from the AAL set (Tzourio-Mazoyer, 2002) from MarsBar and the whole brain analysis. In the ROI analysis, the caudate nucleus appeared significantly more active when overcoming LI than when succumbing to LI in the left hemisphere, and just below threshold in the right hemisphere. The anterior cingulate cortex (ACC) showed a significant main effect of LI conditions minus PE conditions bilaterally, and a sub threshold trend ($p=0.06$ 1 tailed) towards increased activity for overcoming LI compared with succumbing to LI (see: Fig. 8.7). This latter result was also found using a whole brain analysis, although it did not stand up to correction. However, Fig. 8.8 shows the contrast of the two LI conditions at $p=0.005$ uncorrected with a voxel extent threshold of 50 with the main cluster centred near the ACC. Unfortunately, no other significant activations were observed for any contrasts involving the LI and PE task comparisons, even though behaviourally the test was
sensitive enough to reveal differences between these conditions at the behavioural level. This lack of significant results paralleled by significant behavioural differences suggests that the task was probably underpowered because of the noisy nature of fMRI, and the experiment needs to be repeated with a higher number of events. Moreover, given the limits of spatial resolution of fMRI with human subjects, one can assume that these results are broadly consistent with those from the animal literature that have demonstrated the involvement in LI of various components of the hippocampal formation (Gray et al., 1995, 1997; Weiner & Feldon, 1997).

The pattern of results obtained by Grusza, Hampshire and Owen (in prep.) appears to indicate that the ACC and caudate may play roles in dealing with/overcoming LI and not PE. The role of ACC and caudate in executive control is well established, also the effect of PD on these areas is known. However, the result is unexpected given previous findings and given the lack of DA modulation on LI. Further research is required to confirm the relationship of this very preliminary result to previous behavioural studies of LI. More specifically, in order to determine if the repeated overcoming of LI recruits the ACC and caudate, the reliability of the result needs to be confirmed in a follow up study that has greater statistical power. Furthermore, the questions must be answered as to whether LI is still impaired in PD over multiple repeated trials and if such an impairment is still evident, whether that impairment is still insensitive to dopaminergic medication.

**Fig. 8.7** Learned irrelevance and perseveration: contrasting learned irrelevance overcoming vs. succumbing conditions at the moment of set change and contrasting learned irrelevance and perseveration at the moment of set change (from Grusza et al., in prep.)
Fig. 8.8 Learned irrelevance: contrasting overcoming vs. succumbing to learned irrelevance. The whole brain analysis at $p=0.005$ uncorrected level but with a voxel extent threshold of 50 (from Gruszka et al., in prep.)

References


Learned Irrelevance Revisited: Pathology Based Individual Differences


