Poor Set-shifting and Weak Coherence as Neurocognitive Endophenotypes of Eating Disorders

Roberts, Marion

Awarding institution:
King's College London

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Poor Set-shifting and Weak Coherence as Neurocognitive Endophenotypes of
Eating Disorders

Marion Elizabeth Roberts

Institute of Psychiatry, Division of Psychological Medicine & Psychiatry, Section of
Eating Disorders, King's College, University of London

Thesis submitted to King's College London for the degree of
Doctor of Philosophy

2009

1   Abstract

With the lack of effective treatment available for eating disorders (ED), novel approaches to the understanding and therefore treatment of these devastating illnesses is imperative. Recently in the field of psychiatry, the search for endophenotypes (underlying traits that are associated with but not a direct symptom of the illness) has received much attention. Cognitive flexibility (set-shifting) and weak coherence (bias toward local processing) are two aspects of neurocognition that have recently been implicated as candidate endophenotypes of ED. This thesis is the first body of work to systematically assess these cognitive styles in ED using four of the endophenotype criteria outlined by Gottesman and Gould (2003): the endophenotype must 1) be present in the illness population; 2) be state-independent; 3) co-aggregate within affected relatives; 4) present in unaffected relatives at a higher rate than the general population.

Participants were 270 women with current or past ED, sister pairs concordant or discordant for ED, and healthy control women. All participants were administered a neuropsychological battery measuring set-shifting and weak coherence along with self-report questionnaires. Clinical participants were additionally assessed with the SCID for lifetime ED pathology and comorbidity.

Across endophenotype criteria there was moderate evidence for poor set-shifting and strong evidence for weak coherence as endophenotypes of ED. Traits were less notable in the recovered population, suggesting a reduced bias with illness recovery. Effect sizes were small between unaffected sisters and control women. Weak coherence presented differently in those with anorexia and their unaffected sisters (superior local processing) compared to bulimia and their unaffected sisters (poor global integration). Analysis of extreme scores showed that 'impaired shifting' was present in 37% of participants with current ED, while 'persistent detail focus' across coherence tasks was found in just under half of cases. Only 15% of those with current ED showed extreme scores across both cognitive styles. These traits were associated with poor prognostic factors.

The implications of these findings are discussed with particular emphasis on clinical applications. Methodological recommendations and future directions for this field of study are presented.

Acknowledgments

Thank you first and foremost to my amazing supervisors Prof. Janet Treasure and Dr. Kate Tchanturia. Janet, your limitless creativity and devotion to your work have been and will continue to be an inspiration. Thank you for opening the door for a Kiwi girl, for all your support, generosity, and 4am emails ☺. Kate, your neuropsychological expertise and systematic approach to clinical research, all delivered with a loving heart, has been invaluable to me. Thank you for your consistent support, guidance, encouragement and friendship. Thank you both for always having my future in mind, not just this one season.

Many thanks to my colleagues and friends at Guy's and the IoP, working with you all has made this journey a pleasure. To Profs Ulrike Schmidt and Iain Campbell. Special thanks to Carolina, my little Chilean friend, for our shared data discussions and conference hotel rooms! Also special thanks to Frankie (recruiter extraordinaire!) for your energy, persistence and friendship. To those who went before me: Ana-Rosa, Olivia, Fabrice, Jenna, Laura, & Emma, and to those following: Amy, Anna (thanks for the proof-reading!), Liz, Simone, and the rest of the netball girls (go peas!). Thank you all for making the unit such a supportive, collaborative adventure ☺. To the volunteers and students who contributed to this project, especially Betsy, Lucy and Emily, thanks so much for your hard work. To Ellie, our colourful inspiration, for being both a joy and a tangible reminder of the reality of an eating disorder.

A huge thank you to the 270 women who literally gave their blood, and some their sweat and tears to make this research possible. Your contribution has been invaluable. May this thesis help facilitate the provision of targeted and effective treatment that so many of you have not received.

Thank you to the Nina Jackson Research Into Eating Disorders (RIED) trust, in conjunction with the Psychiatry Research Trust (registered charity no. 284286), for awarding me a studentship that enabled this project to be completed.

Thank you to my Parents and Grandparents, whose decisions in life have given me both the solid foundation and the opportunity to do what I love. Thank you for your limitless support, generosity, and love. To my church family and friends in London and Wellington, thanks for listening when I was going slightly mad, for your encouragement, your prayers, and our combined investment in Starbucks ☺.

To my husband Dave, my constant supporter and encourager, thank you for your natural assumption of my success and for boasting about it on the way. Love you more than words can say (even though you haven't read this yet!).

And to Father God, thank you for my passion for this journey, thank you for the experiences, the blessings and the provision that has lead me to this point. May it continue for the purposes of the kingdom, Amen!

Abbreviations used throughout this thesis:

AN	Anorexia Nervosa
ANBP	Binge/purging type Anorexia Nervosa
ANR	Restricting type Anorexia Nervosa
ANOVA	Analysis of Variance
ASD	Autistic Spectrum Disorder
BDD	Body Dysmorphic Disorder
BMI	Body Mass Index
BN	Bulimia Nervosa
CRT	Cognitive Remediation Therapy
DSM-IV	Diagnostic and Statistical Manuel of Mental Disorders, 4th Edition
ED	Eating Disorder or Eating Disorders
EDNOS	Eating Disorder not otherwise specified
EFT	Embedded Figure Test
ES	Effect Size
GAD	Generalised Anxiety Disorder
GEFT	Group Embedded Figure Test
HADS	Hospital Anxiety and Depression Scale
HC	Healthy control or Healthy controls
Hsis	Healthy sister or Healthy sisters
KW	Kruskal-Wallis Test
MW	Mann-Whitney U Test
OCD	Obsessive-compulsive Disorder
OCI-R	Obsessive-compulsive Inventory Revised
OCPD	Obsessive-compulsive Personality Disorder
PTSD	Post Traumatic Stress Disorder
rcAN	Recovered Anorexia Nervosa
RCT	Randomised Controlled Trial
ROCF	Rey-Osterrieth Complex Figure
TMT	Trail Making Test
WCST	Wisconsin Card Sorting Test

Table of Contents

Table of Tables

Table of Figures

Table of Appendices

2   Thesis Outline

The overall aim of this thesis is to investigate two specific aspects of neuropsychological profile (poor set-shifting and weak coherence) as endophenotypes of anorexia and bulimia nervosa. This investigation will be systematically conducted, assessing four of the five criteria outlined by Gottesman and Gould (2003) that must be met in order for an underlying trait to be considered a psychiatric endophenotype.

The first chapter (chapter 3) provides a general overview of the current knowledge base for eating disorders (ED), introduces the concept of an endophenotype and the use of neuropychology to measure candidate traits. Chapter 4 presents a systematic review and meta-analysis of set-shifting ability in ED, as background to the empirical studies. The general methodology employed in this thesis is then presented in chapter 5.

The main empirical studies (chapters 6-9) are organised by endophenotype criteria. Chapter 6 investigates the traits of poor set-shifting and weak coherence in women with current ED compared to healthy control women. Chapter 7 investigates these traits in women recovered from AN. Chapter 8 compares sister pairs concordant for an ED, and chapter 9 compares unaffected sisters of women with ED to both control women and their ED sisters. Chapter 10 explicitly investigates the relationship between poor set-shifting and weak coherence. The impact of neuropsychological traits on illness variables is explored for each clinical group within each of these empirical chapters.

To conclude, the final chapter (chapter 11) outlines the main neuropsychological findings by hypotheses and discusses the strengths, weaknesses and limitations of this thesis. Clinical applications of the findings are detailed, along with recommendations for future research following this line of work.

3   Introduction to the eating disorders and endophenotypes

3.1   Introduction to the eating disorders[1]

Anorexia (AN) and bulimia nervosa (BN) together affect between 4.8 and 12.5% (lifetime, depending on DSM-IV diagnoses included) of the female population at a clinical level (Wade, Bergin, Tiggemann, Bulik, & Fairburn, 2006). Community studies reveal sub-threshold eating disorder (ED) behaviour in up to 33% of female university students (Nelson, Hughes, Katz, & Searight, 1999; Roberts, 2006). While they are often glamorised in the media, ED have the highest mortality rate of any psychiatric condition, and a suicide rate 200 times that of the normal population (Harris & Barraclough, 1998). Both AN and BN are serious, often long-term psychological illnesses whose effects can be disastrous not only for the sufferer, but also for their family and friends (Treasure, Murphy, Szmukler, Todd, Gavan, & Joyce, 2001; Whitney, Murray, Gavan, Todd, Whitaker, & Treasure, 2005). By depriving themselves of the food their body needs to survive, the threat of death can be a reality for those with AN. Prolonged BN can lead to irreversible damage to the digestive system due to the abuse of laxatives, or continued self-induced vomiting. Of all medical conditions, ED are ranked 15th in terms of 'life lived with disability' (Mathers, Vos, Stevenson, & Begg, 2000).

This introductory chapter will briefly outline the current knowledge base regarding ED diagnostic categories, incidence, risk and maintaining factors, and evidence-based approaches to treatment.

3.1.1   *Diagnostic categories*

The Diagnostic and Statistical Manual of Mental Health Disorders (DSM-IV-TR, APA, 2000) currently distinguishes between three main ED diagnoses: AN, BN and Eating Disorder not otherwise specified (EDNOS). While many patients transition between diagnostic categories (particularly with a long duration of illness), AN and BN remain distinct categories from a diagnostic point of view. Different subtypes make up both AN (restricting or binge/purging type) and BN (purging or nonpurging type).

---

[1] Please note that a version of this chapter has been published:
Roberts, M. E., & Treasure, J. (2008). Eating Disorders. In C. Jackson, K. Hill & P. Lavis (Eds.), *Child and Adolescent Mental Health Today: A Handbook* (pp. 201-213). Brighton, UK: Pavillion House.

3.1.1.1 Anorexia Nervosa (AN)

Perhaps the most defining characteristic of AN is the low weight of the sufferer (Body Mass Index [BMI; kg/m^2] < 17.5), as excessive food restriction results in the obvious state of being considerably underweight. People with AN often develop strict rules and rituals around eating, such as preparation of food, order in which food is eaten, and which cutlery may be used. While some sufferers simply restrict their food (ANR), others also engage in the inappropriate compensatory behaviours (ICB's) seen in BN such as self-induced vomiting and laxative abuse (ANP). Less often, the binge/purge subtype of AN is observed (ANBP) where binge episodes are usually less frequent and milder in terms of food consumption (allowing the patient to maintain their low weight), and are coupled with ICB's. While these three subtypes are often used for research, the DSM-IV specifies two subtypes; ANR and ANBP (ANP + ANBP).

In general, the personality type of those with AN is relatively consistent. While both ED share high rates of depression (Godart et al., 2007), people with AN tend to be high academic achievers, displaying perfectionstic tendencies (Bardone-Cone et al., 2007) that are often present from childhood (Brecelj-Anderluh, Tchanturia, Rabe-Hesketh, & Treasure, 2003). High rates of anxiety disorders, particularly obsessive-compulsive disorder and social phobia, are also seen in AN (Kaye, Bulik, Thornton, Barbarich, & Masters, 2004; Godart et al., 2006).

3.1.1.2 Bulimia Nervosa (BN)

Though it is the more common of the ED, the normal weight of BN often makes it difficult to identify. A normal weight is sustained through regular binge eating episodes, where the individual feels distressed and out of control with the large amount of food they are consuming. Binges are compensated for by purging (e.g. self-induced vomiting, laxatives, diet pills, ipecac medication), or less commonly by using non-purging methods such as excessive exercise or dietary restriction. In most cases, people with BN seem to have normal eating patterns to those around them.

In addition to their eating problems, women with BN often present with notable impulsive behaviours, as evidenced by the high rate of alcohol and drug use (Dunn, Larimer, & Neighbors, 2002; Gadalla & Piran, 2007), self-harm (Favaro et al., 2007), and borderline personality disorder (Masjuan, Aranda, & Raich, 2003). Those with BN also display higher levels of posttraumatic stress disorder than those

with AN (Kaye et al., 2004), indicating that their lives are perhaps more chaotic and traumatic.

3.1.1.3 Eating disorder not otherwise specified (EDNOS)

The category of EDNOS is employed when a clinically significant ED is present, however not all of the criteria are met for AN or BN as outlined in the DSM-IV (APA, 2000). Up to 42% of cases presenting to community ED clinics fall into this 'leftovers' category (Button, Benson, Nollett, & Palmer, 2005). Binge eating disorder (BED), a condition where an individual engages in binge eating without any ICB's, is currently classified under the EDNOS umbrella however growing research evidence suggests that BED may be clinically useful as an official diagnosis (Ackard, Fulkerson, & Neumark-Sztainer, 2007). Other examples of EDNOS include someone of normal weight who engages in purging such as vomiting or laxative use in the absence of binging episodes ('purging disorder'), or someone with AN whose weight is not low enough to merit full diagnosis (i.e. periods remain regular). This difficulty with classification is especially obvious in the child and adolescent population, prompting some researchers to develop modified criteria for children and adolescents with ED (Nicholls, Chater, & Lask, 2000).

It is difficult to discuss those with EDNOS as a category, given that it represents such a mixed bag of disordered eating behaviours. This thesis will focus on the two main categories of ED; AN and BN.

3.1.2 Incidence

The often quoted statistic for the incidence of ED among western females is 0.3% lifetime for AN, and 1% lifetime for BN (Hoek & van Hoeken, 2003). However research findings indicate that ED can affect up to 12.5% of the female population at a clinical level (Wade et al., 2006), and up to a third of female university students show disordered eating behaviour (Nelson et al., 1999; Roberts, 2006). A systematic review of the adolescent literature notes that many adolescents suffer from partial, or sub-threshold ED (EDNOS) and that while they do not reach full diagnostic criteria, these individuals still suffer to a significant degree (Chamay-Weber, Narring, & Michaud, 2005).

AN was first described as an illness by Sir William Gull, a physician at Guy's Hospital London in the mid 1800's (Gull, 1868). The incidence of AN increased substantially in the mid 20th century levelling off in the 1970's (Schmidt, 2005), however the age of sufferers has continued to drop with girls as young as 6 years old

being admitted for treatment (Nicolls, 2007). BN is by comparison a recent condition, first entering the DSM-III in the 1980's after being described by Prof. Gerald Russell (Russell, 1979). This official classification as a psychiatric condition saw an unsurprising increase in the prevalence of BN, which has only recently levelled off (Currin, Schmidt, Treasure, & Jick, 2005). The volume of highly palatable, calorific foods available in our culture over the last few decades has allowed for easy access to "binge" foods, that simply would not have been accessible 50 years ago, thus increasing the viability of bulimic behaviour. This availability of food likely contributes to the incidence of BN being five times higher in cities than rural areas, whereas no difference in urbanisation is found regarding incidence of AN (van Son, van Hoeken, Bartelds, van Furth, & Hoek, 2006).

A substantial gender disparity is observed in the ED, with the prevalence of male ED approximately 5-10% of that of females (Schmidt, 2005). A review of 813 cases of AN referred to a specialist ED clinic found that nearly all aspects of eating disorder pathology were the same across the two sexes (Crisp et al., 2006). Males accounted for 7.6% of this AN cohort, in keeping with the incidence rate above.

3.1.3   Risk and Maintaining factors

Both AN and BN are the result of a complex interaction of multiple factors; cultural, environmental, psychological, and biological; which are yet to be fully understood. Perhaps the most obvious risk and maintaining factor for an ED from a lay perspective is the social climate in which we currently live. Young women and models alike are faced with more pressure than ever before to be thin, with messages from the media and the catwalk constantly pushing unrealistic and unhealthy body images (Treasure, Wack, & Roberts, 2008). These pressures are understandably predictive of eating pathology (Stice, 2004), and are felt in girls as young as age 5 (Dohnt & Tiggemann, 2006). However, it is important to keep the influence of culture in the context of other risk factors. AN and BN share many risk factors such as gender, ethnicity, genetics, childhood anxieties, aversive life events, acculturation, and negative self-evaluation (Jacobi, Hayward, de Zwaan, Kraemer, & Agras, 2004). However a number of risk factors exclusive to each disorder have been identified.

3.1.3.1   Risk factors in Anorexia and Bulimia Nervosa

Following an extensive review of the literature, Jacobi and colleagues have outlined specific risk factors, weighted by potency, for both ED. Specific to AN are obsessive-compulsive personality traits throughout childhood and adolescence

(medium potency), perfectionism in late adolescence (medium potency), and high level of exercise at approximately age 13 (high potency). In addition to environmental factors, understanding the genetic or heritable aspect of AN has taken a crucial step forward in the last decade of research. ED are often found to run in families, with the relative risk of developing AN at 11.2 for female relatives of those with AN, and 12.3 for female relatives of those with BN (Schmidt, 2005). Susceptibility genes for AN such as brain-derived neurotropic factor (BDNF), along with the serotonin system (5-HT), have been identified in molecular genetics (Collier & Treasure, 2004). These relatively new advances in understanding of the genetic risk of AN require further exploration and replication.

Like AN, a number of specific risk factors have been identified for BN. Childhood obesity from as early as age 4, along with parental obesity are medium potency risk factors (Jacobi et al., 2004). Dieting behaviour in mid adolescence is, perhaps surprisingly, a high potency risk factor for BN rather than AN. Additional parental factors such as parental alcohol and drug use, parental depression, parental criticism (high expectations, comments on weight, low contact with the adolescent) and adverse family experiences factor as specific risks for BN. While genetic advances in BN are being made alongside those for AN, it seems the former is to a lesser extent influenced by genes. The relative risk for developing BN is considerably lower than AN, reported at 4.2 for female relatives of those with AN, and 4.4 for female relatives of those with BN (Schmidt, 2005).

3.1.3.2   *Maintaining factors for Anorexia Nervosa*

A number of theories have been put forth over the past few decades, in an attempt to explain the long duration of illness observed in the ED. Schmidt and Treasure (2006) recently outlined a model of maintaining factors in AN, that avoids cultural aspects of the disorder such as weight and shape related pressures. Four factors are proposed:

1) Perfectionism/cognitive rigidity/obsessive-compulsive traits: (often seen in childhood, also reflected in the high rate of co-occurring obsessive-compulsive disorder and obsessive-compulsive personality disorder in anorexia);

2) Experiential avoidance (essentially avoiding emotion and emotional memories, which is aided by the maintenance of a low weight);

3) Pro-anorectic beliefs (positive thoughts about the value of the AN);

4) Response of close others.

The model proposes that these four factors interact to encourage the patient to maintain the AN behaviours. The first two factors are often present prior to (and are exacerbated by) the illness, while the second two develop in response to the AN. It is proposed that by addressing each of these points in specialist therapy, a more positive outcome will be seen. The current thesis will investigate traits based on factor one of this model.

3.1.4   Evidence-based psychological treatment

It is widely recognised in the field that the earlier someone with an ED is identified and provided with appropriate treatment, the greater their chance of a quick and full recovery. However a barrier to delivering speedy treatment is that many patients (particularly those with AN) consider their ED to be a solution rather than a problem, and their abnormal eating patterns are often kept secret. Motivation to change is therefore a key issue in treatment. The obvious life endangering signs of AN can cause family members to intuitively protect the affected individual, often creating stress in the household and tension between the families' and patients' readiness to change.

Recent systematic reviews of the treatment research in eating disorders reveal limited numbers of Randomised Controlled Trials (RCT's) investigating treatment outcome for AN, BN, and BED (Brownley, Berkman, Sedway, Lohr, & Bulik, 2007; Bulik, Berkman, Brownley, Sedway, & Lohr, 2007a; Shapiro, Berkman, Brownley, Sedway, Lohr, & Bulik, 2007). None of these address medication treatment.

3.1.4.1   Treatment for Anorexia Nervosa

As can be seen in Table 1, only 16 randomised controlled trials (RCTs) for AN have been undertaken over the last two decades. Four of these were completed subsequent to the NICE guidelines being published. One trial completed since the NICE guidelines questioned whether it was possible to use such a methodological approach for this condition, as the acceptability of some treatments was so low (Halmi et al., 2005). In addition to this paucity of data, inconsistent styles of reporting outcome variables make it difficult to generalise findings across studies.

Family Therapy has been recognised as the most promising form of therapy for children and adolescents with AN, as recommended by the NICE guidelines (Grade B, 2004). A manual describing this style of family therapy, the so-called "Maudsley Model", has been produced by James Lock and colleagues (Lock, Grange, Agras, & Dare, 2001). A number of RCTs have compared various 'doses'

Table 1: Outline of Randomised Controlled Trials in the Anorexia Nervosa Literature

RCT	N	Age (SD)	Baseline BMI/IBW	Intervention	Pre/Post BMI/IBW Effect Size	Treatment Completers	"Good" Outcome
Gowers et al. (2007)	167 (AN)	14.9	15.3 (1.6)	In-patient Individual and Family Therapy (n=57)	1.59	49.1%	26.3%
			15.3 (1.6)	Specialised out-patient (n=55)	1.80	74.5%	47.3%
			15.5 (1.6)	CAMHS treatment as usual (n=55)	1.79	69.1%	47.4%
Halmi et al. (2005)	122 (AN)	24.8 (6.8)	17.8 (1.7)	Cognitive Behavioural Therapy	-	43.0%	-
				Fluoxetine	-	27.0%	-
				CBT + Fluoxetine	-	41.0%	-
Lock et al. (2005)	86 (AN)	12-18yrs	17.1 (1.4)	Short-term Family Therapy (n=44)	1.4	95.4%	-
				Long-term Family Therapy (n=42)	1.22	83.3%	-
McIntosh et al. (2005)	56 (AN)	17-40	17.3 (1.1)	Cognitive Behavioural Therapy (n=19)	0.6	63.0%	-
				Interpersonal Psychotherapy (n=21)	0.44	57.0%	-
				Specialist Supportive Clinical Management (n=16)	1.1	69.0%	-
Ball & Mitchell (2004)	16 (ANR)	18.01 (2.97)	16.26	Cognitive Behavioural Therapy (n=9)	1.74	69.0%	60.0%
	9 (AN-BP)			Behavioural Family Therapy (n=9)	1.78	75.0%	(Total)
Pike et al. (2003)	33 (AN)	26.1 (6.2)	16.0 (2.1)	Cognitive Behavioural Therapy	-	77.8%	44.0%
		24.3 (6.9)	15.2 (1.5)	Nutritional Counseling (n=	-	26.7%	7.0%
Dare et al. (2001)	84 (AN)	26.3 (6.7)	15.4 (1.6)	Focal Psychotherapy (n=21)	-	57.1%	33.3%
				Family Therapy (n=21)	-	76.2%	36.4%
				Cognitive-analytic Therapy (n=21)	-	62.0%	9.0%
				Routine' Treatment (n=21)	-	62.0%	5.3%
Eisler & Dare (2000)	40 (AN)	15.5 (1.6)	74.3% (9.8)	Conjoint Family Therapy	-	85.0%	26.0%
				Separated Family Therapy	-	75.0%	47.6%
Geist et al. (2000)	25 (AN)	14.3 (1.5)	74.9 (9.2)	Family Therapy (n=12)	2.94	-	-
		14.9 (1.7)	77.2 (11.1)	Family Group Psychoeducation (n=13)	2.68	-	-
Wallin et al. (2000)	23 (ANR)	NR	15.45 (1.75)	Body Awareness Therapy + Family Therapy	-	-	61.5%
	3 (AN-BP)			Family Therapy	-	-	69.2%
Robin et al. (1999)	37 (AN)	14.5	15.9	Behavioural Family Systems Therapy (n=19)	2.61	-	-
				Ego-oriented Individual Therapy (n=18)	0.9	-	-
Serfaty et al. (1999)	25 (AN)	22.1 (6.6)	16.6	Cognitive Therapy (n=25)	0.81	92.0%	-
				Dietary Advice (n=10)	N/A	0.0%	-

Gowers et al. (1994)	40 (AN)	21.2 (5.12)	15.52 (1.44)	Outpatient Psychotherapy (n=20)	2.05	75.0%	60.0%
		21.9 (4.460	15.84 (1.67)	Control- One off assessment (n=20)	0.48	N/A	-
le Grange et al. (1992)	18 (AN)	15.33 (1.81)	77.9 (7.62)	Family Therapy- conjoint (n=9)	1.23	-	-
				Family Counseling- separated (n=9)	2.83	-	-
Channon et al. (1989)	24 (AN)	23.8 (6.28)	15.3	Cognitive Behavioural Therapy (n=8)	-	100.0%	-
				Behavioural Therapy (n=8)	-	87.5%	-
				Control (n=8)	-	75.0%	-
Russell et al. (1987)	57 (AN)	21.8 (7.1)	69.6 (13.0)	Family Therapy (n=36)	-	63.4%	22.0%
	23 (BN)			Individual Therapy (n=37)	-	68.4%	16.2%

RCT Randomised Controlled Trial; AN Anorexia Nervosa; ANR Restricting type Anorexia Nervosa; ANBP Binge/purge type Anorexia Nervosa; BN Bulimia Nervosa; BMI Body Mass Index; IBW Ideal Body Weight

- Data not reported, or insufficient data in paper to allow calculation

of family therapy for adolescents (le Grange et al., 1992; Robin et al., 1999; Eisler et al., 2000; Geist et al., 2000; Lock et al., 2005), with small outcome differences between these variants being observed. Large gains in BMI (as measured by effect size) for the groups receiving family therapy are seen across all trials, and treatment completion is high. A recent RCT compared specialised inpatient, multimodal specialised outpatient (CBT, motivation, feedback, parental involvement, dietary advice) and general outpatient treatment for children and adolescents with AN (Gowers et al., 2007). Little benefit was found for inpatient management, where adherence to treatment was less than 50%.

There is less certainty about treatment for adults. The overall outcome is worse and there is little difference between the types of therapy (family or individual, CBT or dynamic). One surprising finding was that a non-specific supportive form of treatment out performed CBT (McIntosh et al., 2005). If replicated, this may have important repercussions in terms of understanding the process of change in AN as so far other treatment trials have little difference between more specialist treatments. There is some evidence that routine psychiatric treatment has a poorer outcome (Dare et al., 2001) so these results are not merely a placebo response. A key problem with these limited studies is that such treatments are simply being borrowed from treatments developed for other disorders, rather than specifically developed for the purpose of treating AN.

3.1.4.2   Treatment for Bulimia Nervosa

A large body of evidence exists in favour of Cognitive Behavioural Therapy (CBT) as the treatment of choice for BN (Shapiro et al., 2007), leading the NICE guidelines to recommend CBT as a grade A treatment for adult BN. Due to the lack of adolescent specific treatment trials available at the time, NICE guidelines also recommended CBT for adolescent BN, adapted to consider the younger age of the patients and to include family members whenever possible. Since then, attempts have been made to expand research in this area (Le Grange & Schmidt, 2005), with two case series' indicating the potential for tailored CBT in the adolescent population (Lock, 2005; Schapman-Williams, Lock, & Couturier, 2006). A Maudsley-based family approach has also since been developed (le Grange & Lock, 2007). The first RCT of adolescent BN, published by Schmidt and colleagues, found a CBT guided self-care (supported by a professional) was more effective at 6 month follow-up than family therapy (Schmidt et al., 2007). This difference disappeared at 12 month

follow-up however given the smaller direct cost of treatment for the self-care group this is an exciting finding.

3.2   Introduction to endophenotypes in psychiatry

Psychiatric illnesses are complex disorders both in terms of aetiology and presentation, making research based on overt clinical symptoms particularly challenging. An alternative strategy is to focus on more primary causative markers. An endophenotype refers to an underlying cognitive or behavioural trait that is associated with a disorder, but is not part of its visible presentation or 'phenotype'. Thus the endophenotype falls on the pathway between abnormal behaviour and biology (phenotype and genotype), and indeed is often referred to as an intermediate phenotype or biological marker (biomarker). The concept of an endophenotype was first introduced in the schizophrenia field in 1973 (Gottesman & Shields, 1973). It was proposed that moving focus from the diverse phenotypes of schizophrenia to endophenotypes would provide a simpler architecture with which to identify the genetic basis of psychiatric illness, as compared to diverse and inconsistent illness symptoms. For example, in schizophrenia hallucinations and/or delusions are prominent phenotypes of the illness and indeed form part of the diagnostic criteria. Fixed marker research based on the presence or absence of a particular type of delusion can become immediately irrelevant with a change in illness presentation. However a trait like poor working memory, whilst not part of diagnostic criteria, often exists alongside the more obvious schizophrenic traits and may represent a stable underlying component of the illness. Using this example, poor working memory may provide a more direct association to the genetic architecture of schizophrenia than do hallucinations or delusions.

In a seminal paper by Gottesman and Gould (2003), five criteria for empirically assessing an endophenotype were outlined:

(1) The endophenotype is associated with illness in the population.

(2) The endophenotype is heritable.

(3) The endophenotype is primarily state-independent (i.e. present in acute and recovered phases of the illness).

(4) Within families, the endophenotype and illness co-segregate.

(5) The endophenotype found in affected family members is found in non-affected family members at a higher rate than in the general population.

Also detailed were additional descriptors of candidate endophenotypes, in that they may be neurophysiological, biochemical, endocrinological, neuroanatomical, cognitive, or neuropsychological in nature (Gottesman & Gould, 2003). Flint and Munafo extended and refined these categories, listing anatomical (e.g. decreased grey matter in schizophrenia), developmental (e.g. age at first word in autism), electrophysiological (e.g. exploratory eye movement in schizophrenia), metabolic (e.g. cortisol secretion in anxiety), psychological (e.g. cognitive function in bipolar disorder and ADHD) and sensory deficits (e.g. olfactory sensitivity in schizophrenia) as candidate endophenotype categories (Flint & Munafo, 2007). Further suggested refinements to the criteria include the endophenotype being linked to the causal process (Walters & Owen, 2007), involved in plausible biological mechanisms, predictive of the disorder probabilistically, and lying closer to the site of the primary causative agent (Flint & Munafo, 2007). The primary criteria of Gottesman and Gould (2003) provided a timely framework within which investigations of candidate endophenotypes across the diverse fields of mental health could be systematically measured.

While the 30 years following the introduction of the endophenotype concept saw only 37 papers published (PubMed search 'endophenotype' 1973-2003, search conducted 9 July 2008), the last five years has seen a relative explosion in endophenotype research with over 360 articles in the literature. This may in part be due to the availability of more sophisticated genetic methods to measure heritability of candidate endophenotypes. Of more relevance to the current thesis, psychiatry is becoming more aware of the lack of biological basis of current diagnostic classification systems. Perhaps the biggest contribution of endophenotype research aside from furthering genetic understanding is the potential to develop an empirically based biological framework for mental illness, which can be used as a foundation for more biologically relevant classification systems. Classification modelled in such a way may provide a more accurate descriptor of the different presentations of ED, which in turn would provide more precise information about illness characteristics, likely prognostic factors, and response to various forms of treatment.

3.2.1 *Endophenotype criteria in this thesis*

This thesis will use the endophenotype criteria outlined by Gottesman and Gould (2003) to assess two candidate endophenotypes of ED, namely poor set-shifting and weak coherence. The large sample size required for DNA analysis and

therefore assessment of the heritability criteria for an endophenotype (criterion 2) was outside the scope of the current study. This leaves the four 'behavioural' criteria of an endophenotype which will be systematically addressed in this thesis, with the genetic aspect forming part of a larger ongoing project.

Gottesman and Gould (2003) use criterion 2 to distinguish between an endophenotype (where there is evidence of heritability) and a biomarker (where there is no evidence of heritability). As heritability will not be directly addressed here it could be argued that this thesis would more appropriately be entitled an assessment of candidate biomarkers than an assessment of candidate endophenotypes. However given that data for genetic analysis is collected for future analysis and that no evidence against heritability for the concepts under question is found, this thesis will use the term endophenotype rather than biomarker. For clarities sake, each of the four behavioural endophenotype criteria will be addressed in the sequence outlined by Gottesman and Gould, numbered 1-4.

3.3   Using neuropsychology to measure candidate endophenotypes

As mentioned above (3.2), one medium used to explore candidate endophenotypes is that of neuropsychology, a field with its roots in the observation of behaviour change following traumatic brain injury. One such seminal patient is that of Phineas Gage who, following a steel rod puncturing through his frontal lobes in 1848, lost the ability to successfully control and regulate his behaviour however much to the surprise of physicians at the time, remained otherwise in good health (Harlow, 1848; Damasio, Grabowski, Frank, Galaburda, & Damasio, 1994). Half a century later, high rates of brain injury during the First World War prompted the need for standardised assessment of cognitive functioning to inform rehabilitation (see references in Lezak, Howieson, Loring, Hannay, & Fischer, 2004). These early attempts were further refined into more sophisticated assessment and treatment techniques for veterans of World War II. In parallel to these advances in medicine, psychologists were attempting to develop measures of the concept of intelligence pioneered by Binet and colleagues in France. The Stanford-Binet intelligence test (circa 1916) was one of the first attempts at a task battery in order to conduct large-scale screening of individuals for intelligence; the latest revision of which measures verbal and non-verbal intelligence across five domains such as visuo-spatial processing and working memory (Becker, 2003). The early work of Binet lead to the

development of various batteries many of which are still used today, such as the Wechsler Intelligence Scales, and Raven's Progressive Matrices. The development of these tests for screening and educational purposes allowed the field of psychology to understand cognitive processes from a normative viewpoint (Lezak et al., 2004). Experimental studies of cognitive functioning in humans and animals have also contributed to the development of neuropsychological assessment. Profiling of both normal (usually student) populations in addition to those with localised brain injury such as stroke patients or those with traumatic brain injury has contributed to our understanding of how cognitive processes can become altered or impaired.

On applying neuropsychological assessment to the psychiatric population, differences in cognitive profile are notably subtler. This is because deficits are not caused by explicit damage such as stroke or brain injury, but rather represent more delicate neurological abnormalities that are now understood to underpin many forms of psychiatric illness for example schizophrenia and depression (Lezak et al., 2004). Neuropsychological assessment allows for more detailed understanding of altered cognitive functioning in the psychiatric population.

3.3.1   Cognitive functions as candidate endophenotypes

Executive functioning refers to "those capacities that enable a person to engage successfully in independent, purposeful, self-serving behaviour" (Lezak et al., 2004). The executive functions cover a wide range of skills and do not represent a unitary concept, as evidenced by the dissociation across tasks often observed clinically in individuals with brain injury. Miyake et al. (2000) examined three key aspects of executive functioning to assess the unity of the concept; cognitive set-shifting, information updating and monitoring (working memory), and response inhibition. It was found that while the three aspects were moderately correlated with each other, structural equation modelling indicated that they remained distinct contributors to overall executive functioning.

Deficits in executive function processes are not only found in those with traumatic brain injury. Impaired executive functioning has been implicated as a key aspect of psychiatric illness particularly in schizophrenia (Reichenberg & Harvey, 2007) and ADHD (Doyle, 2006), where deficits in working memory, inhibition of response, and shifting aspects have been highlighted. For example, deficits in cognitive set-shifting have been found in children with ADHD (Halperin, Trampush, Miller, Marks, & Newcorn, 2008) and children with OCD (Shin, Choi, Kim, Hwang,

Kim, & Cho, 2008). A gender disparity in spatial working memory deficit has been found in those with bipolar disorder, where contrary to the pattern in the general population, males compared to females with bipolar disorder show poorer working memory strategy (Barrett, Kelly, Bell, & King, 2008). Additionally, those with schizophrenia have been found to show a more pronounced deficit in shifting set than those with bipolar disorder, who were also impaired compared to healthy controls (Wobrock, Ecker, Scherk, Schneider-Axmann, Falkai, & Gruber, 2008). Various deficits in executive functioning have been identified in the ED literature including working memory (Kemps, Tiggemann, Wade, Ben-Tovim, & Breyer, 2006) and response inhibition (Southgate, 2005) or it's opposite, impulsivity in the BN disorders (Rosval, Steiger, Bruce, Israel, Richardson, & Aubut, 2006).

3.4 Candidate endophenotypes of eating disorders

The concept of an endophenotype can be aptly applied to any mental illness, particularly ED given their unstable phenotypic diagnostic categories. Treatment to date for AN has changed little from when the illness was first described, where focus is placed on remediating the salient phenotype of a refusal to eat (Treasure, 2007). The limited success of this approach (as evidenced by poor response to standardised treatment in both AN and BN) illustrates the need for re-feeding to be coupled with treatment addressing underlying traits and biological causes of the illness (Treasure, Lopez, & Roberts, 2007). Therefore the search for endophenotypes of ED has marked implications for the development of tailored, more effective treatment programs for this resistant population.

The ED field lags behind the rest of psychiatry with regard to research on candidate endophenotypes. Four reviews on the potential of endophenotypes in the ED field are found (Bulik et al., 2007b; Steiger & Bruce, 2007; Treasure et al., 2007; Treasure, 2007), where a number of theoretical candidate endophenotypes are outlined. Impulsivity/reward sensitivity, fear/punishment/stress and social cognition have been suggested as putative endophenotypes (Treasure et al., 2007), along with temperament (Bulik et al., 2007b), reduced serotonin transported activity and novelty seeking (Steiger & Bruce, 2007). However only a handful of empirical studies are as yet available, with contradictory findings already evident. In terms of neurophysiology, serotonin (or 5-HT) has been implicated as a candidate endophenotype due to elevated platelet levels in BN probands and their mothers and

sisters compared to controls using a proxy for central 5-HT reuptake activity (Steiger et al., 2006). However investigation of 5-HT in a different laboratory using single photon emission tomography in discordant BN twins and controls found little difference between groups (Koskela et al., 2007). Methodological differences likely contributed to these mixed findings.

Two candidate neuropsychological endophenotypes have been identified in our group; cognitive set-shifting and weak coherence (discussed below in 3.3). To date, two or three of the criteria for an endophenotype have been addressed for these aspects of neurocognition (Tchanturia, Morris, Anderluh, Collier, Nikolaou, & Treasure, 2004b; Holliday, Tchanturia, Landau, Collier, & Treasure, 2005; Lopez, Tchanturia, Stahl, & Treasure, 2008e). However other than these few studies, no formal investigation addressing all criteria of an endophenotype simultaneously are found in the literature.

3.4.1 *Set-shifting as a candidate endophenotype*

One focus of this thesis will be on cognitive flexibility or set-shifting ability. The neuropsychological literature has distinguished between two aspects of cognitive flexibility: spontaneous flexibility and reactive flexibility (Eslinger & Grattan, 1993). Spontaneous flexibility refers to the "ready flow" of novel responses and ideas, such as the production of multiple different answers in a given situation or in response to a given question. This is referred to as a fluency in thinking, as the standard or habitual response to a given stimuli must be integrated with other seemingly irrelevant knowledge and information, in order for creative generation of ideas to seamlessly occur. This aspect of cognitive flexibility is often tapped with tasks such as verbal fluency (FAS), where patients are asked to generate as many words as they can beginning with F, then beginning with A, then S. Another task used to measure this concept is the uses of common objects task, where the patient is shown a picture of an everyday item (e.g. tin can, cardboard box, book) and asked to generate as many different uses of each item as they can. Such tasks exemplify that spontaneous flexibility is a self-generated, internal process. Difficulties with spontaneous flexibility would manifest as an inability to generate multiple novel responses, perhaps struggling to generate more than one response, or resulting in perseverations of the same response (e.g. on the FAS task a participant may list Fish, Furniture, Fish… Family…. Fish).

In contrast, reactive flexibility refers to the ability to effortlessly change between multiple cognitive sets or behaviours in response to new task demands or changes in situational context. In contrast to spontaneous flexibility, reactive flexibility is initiated by a cue from the external environment, prompting the individual to modify their previous response in favour of a new, now more appropriate response. Perhaps the most common task used to measure reactive flexibility is that of the Wisconsin Card Sorting Test (WCST), where participants sort cards by a given category. After 10 correct sorts, the sorting rule changes and the participant must adapt their response to the new sorting rule to avoid negative feedback. Difficulties with reactive flexibility on this task would manifest as a large number of perseverative responses, where the participant continues to sort based on the previous sorting rule such as colour (despite negative feedback), rather than adapting to the new task demand of sorting by shape. Individuals with impaired reactive flexibility may find themselves to be ritualistic in everyday life, and struggle to cope with unexpected situations, new activities, and changes in plans.

3.4.1.1 *Evidence from general psychiatry*

Much research on both spontaneous and reactive flexibility exists in the psychiatric literature. It is well recognised that cognitive impairment in executive functioning is a core feature of schizophrenia (Wobrock et al., 2008), with strong evidence for cognitive impairment also present in bipolar disorder (Sachs, Schaffer, & Winklbaur, 2007), and obsessive-compulsive disorder where some evidence suggests that impaired flexibility may be state-independent (Rao, Reddy, Kumar, Kandavel, & Chandrashekar, 2008). Table 2 outlines a summary of results from a search of set-shifting ability by psychiatric condition for unaffected 1st degree relatives of those with mental illness. Twenty-five studies were identified in the schizophrenia literature, where the trail making test (TMT), Wisconsin card sorting test (WCST) and/or verbal fluency tasks have been employed amongst twins, siblings and parents of those with schizophrenia. Across all tasks, small to large effect sizes were observed, indicating that set-shifting difficulties are evident in unaffected relatives of those with schizophrenia. Notable variance in effect size across studies was seen on some tasks. For example, effect sizes for WCST perseverative errors were predominantly small to moderate with a maximum effect size of 0.69 (Saoud et al., 2000), where relatives of those with schizophrenia made significantly more errors than controls. However two studies showed a trend in the

Table 2: Summary of set-shifting effect sizes for 1st degree relatives of those with psychiatric illness

	ASD		OCD		BPD		MDD		SZF	
	ES	K	ES	K	ES	K	ES	K	ES	K
TMT-B	0.18	1	0.01	1	0.34	5	0.06	1	0.49	12
TMT B-A	0.34	1	-		-		-		0.59	3
WCST pe	0.31	3	-		0.29	6	0.02	1	0.29	19
WCST cc	-		-		0.02	5	0.07	1	0.37	15
Letter fluency	-0.33	1	-0.12	1	-0.06	5	-		-0.51	11
Category fluency	-0.41	1	-0.17	1	-		-		-0.77	4
IDED errors TC	0.42	2	-		-		-		-	
IDED trials TC	1.03	1	1.01	1	-		-		-	

ASD Autistic Spectrum Disorder; OCD Obsessive-compulsive Disorder; BPD Bipolar Disorder; MDD Major Depressive Disorder; SZF Schizophrenia; ES Effect Size (unweighted mean of study effect sizes- see Appendix 1 for raw data); K n of studies contributing to ES; TMT Trail Making Test; WCST (pe/cc) Wisconsin Card Sorting Test (perseverative errors/categories completed); IDED (TC) Intra-dimensional Extra-dimensional Shift (To Criterion)

opposite direction (Egan et al., 2001; Laurent, Gilvarry, Russel, Mathieu-Cura, & Murray, 2003). This variance contributed to the small effect size for WCST in relatives of those with schizophrenia.

A more consistent profile was seen on the TMT. Results from all studies trended in the direction of rigidity however again with notable variability. On verbal/letter fluency (a task that taps spontaneous flexibility), effect sizes range from negligible (Goldberg et al., 1995) to large (Laurent et al., 1999). At least some of this variance can be attributed to differences in control group performance across studies. For example, the number of raw perseverative errors on the WCST made by control groups ranged from 4.5 (Yurgelun-Todd & Kinney, 1993) to 17.3 (Laurent et al., 2001). In itself, this difference across control samples produces a very large effect size of 1.37 indicating a lack of reliability across control groups. A meta-analytic review of set-shifting in 1st degree relatives of those with schizophrenia was published in 2004 (Sitskoorn, Aleman, Ebisch, Appels, & Kahn, 2004). Like the more up to date data reported here (search conducted October 2008), it was concluded that the cognitive deficits of those with schizophrenia are also found in unaffected 1st degree relatives with a moderate effect size. This result provides evidence for set-shifting as a candidate endophenotype for schizophrenia.

Following schizophrenia, the second largest evidence base on shifting in 1st degree relatives is in the field of bipolar disorder. As illustrated in Table 2, the same tasks have been used in this population however a lesser effect is seen. Negligible to small effect sizes are seen across tasks for bipolar relatives. Again, significant variance between studies is found. On the WCST, effect sizes of relatives compared to controls range from negligible (Kremen, Faraone, Seidman, Pepple, & Tsuang, 1998) to large (Frantom, Allen, & Cross, 2008), as do findings from the TMT. A recent systematic review of neuropsychological deficits in bipolar disorder patients and their 1st degree family members showed effect sizes just short of moderate for TMT and WCST, and a small effect size for FAS fluency (Bora, Yucel, & Pantelis, 2008).

Some evidence for set-shifting impairment in relatives of those with autistic spectrum disorder is found, where like schizophrenia a moderate effect size is found. Differing comparison groups, including healthy controls, parents of children with downs syndrome or mental retardation, and siblings of those with learning disabilities, clouds the autism evidence base. Limited evidence is found in other

disorders, with a small number of familial studies in the depression and OCD literature showing negligible to small effect sizes.

In sum, there is evidence for set-shifting impairment in 1st degree relatives of those with schizophrenia, bipolar disorder and autism where small to moderate effects are seen. This indicates that these psychiatric conditions may share poor set-shifting as an endophenotype. Notable variance is evident across studies. Investigation of poor set-shifting as a familial trait across the broader psychiatric spectrum of disorders such as depression and OCD is poorly developed.

3.4.1.2 Evidence from the eating disorder literature

Reactive flexibility has been studied with great interest in ED over the last decade. Also referred to as "attentional switching" or "task switching", impaired set-shifting in both AN and BN populations is a consistent finding in the literature (Tchanturia, Campbell, Morris, & Treasure, 2005), lending support to criterion 1 of an endophenotype. Poor flexibility has been found to remain in weight recovered and fully recovered individuals with AN (Tchanturia et al., 2004b), lending support to criterion 2 of an endophenotype (primarily state-independent). An investigation of shifting in 42 sister pairs discordant for AN found that like their AN sisters, unaffected sisters were impaired in comparison to control women on the CatBat and Haptic shifting tasks (Holliday et al., 2005), lending support to the final criterion of an endophenotype (present in unaffected 1st degree relatives). These studies require replication.

The current evidence base for impaired set-shifting in both AN and BN will be addressed in more detail in the next chapter ("*A meta-analytic review of set-shifting ability in eating disorders*").

3.4.2 Weak coherence as a candidate endophenotype

Another aspect of cognitive function measured through neuropsychology is that of weak coherence, an information processing style whereby environmental stimuli are processed in a detailed, piecemeal fashion relatively independent of their context. While typically developing individuals tend to integrate information into the 'gestalt', essentially generating the gist from a given situation, those with weak coherence focus on individual details to the extent that their more global significance (or contextual meaning) is often lost. Early conceptualisation of 'weak central coherence' detailed these aspects as linear (see Figure 1) in that a superiority in detail

naturally implied deficits in global integration (Frith, 1989). However more recent theoretical accounts discuss two balanced aspects of weak coherence (see Figure 2); a superiority in detailed processing, and a deficit in global integration (Happe & Booth, 2008). These aspects can but do not necessarily act as a trade-off, in that it is possible to have both or only one aspect of this cognitive style.

3.4.2.1 Evidence from general psychiatry (autism) literature

The weak central coherence account was developed to explain why those with Autistic Spectrum Disorder displayed a tendency to become focussed on small details, and lacked the ability to see these details within their context (Frith, 1989). At the time of its inception, two main theories dominated the autism literature; deficits in executive functioning, and 'Theory of Mind' (the inability explain or predict behaviour from the perspective of others). These theories were thought to best explain the so-called 'triad of impairments' observed in those with autism: impairment in communication; impairment in social skills; and restrictive and repetitive interests/movements/activities. While they appropriately addressed the first two aspects of the triad, executive functioning deficits and theory of mind were unable to satisfactorily explain the non-social impairments and skills of those with autism (e.g. savant abilities, desire for sameness). A preoccupation with details forms part of the diagnostic criteria for autism (APA, 2000) and was indeed not a new concept in the field. Early accounts describing ASD such as those by Kanner detailed an "inability to experience wholes without full attention to the constituent parts" (Kanner, 1943). However the proposal of this often savant-like focus on detail as a theoretical explanation i.e. as an explanation of both the deficits and superior abilities of those with autism was proposed in the literature only 20 years ago.

Since the initial proposal of the weak central coherence account, a variety of neuropsychological tasks have been employed to investigate the concept of weak coherence in the autism population. In a review of evidence to date for and against the weak coherence concept, Happe and Booth (2008) outline numerous studies where evidence of weak coherence is found in autism using a variety of tasks such as the block design task, embedded figures task (EFT), fragmented pictures task, sentence completion task, possible/impossible figures, and drawing tasks such as the Rey-Osterreith complex figure and a free drawing task. They highlight that many tasks simultaneously measure aspects of detail focus and global integration, making it difficult to parse out the individual effect of each processing style. Therefore while

Figure 1: An illustration of the traditional conceptualisation of weak central coherence

Detail focus ———————————————— Global integration

Figure 2: An illustration of the new conceptualisation of weak coherence (Happe & Booth, 2008)

Poor detail focus

Poor global integration ————————————— Strong global integration

Strong detail focus
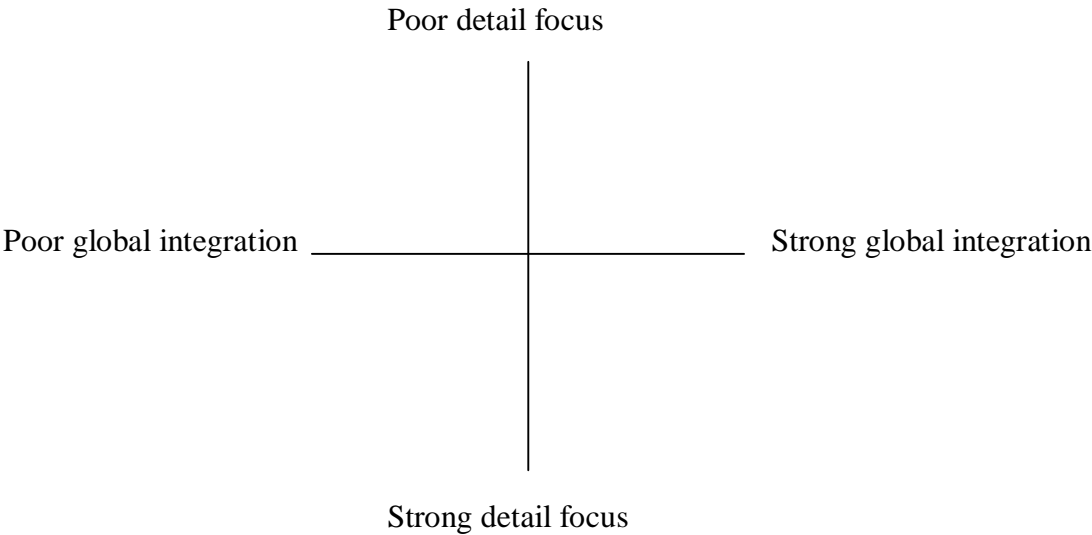
 significant evidence in favour of the weak coherence concept in autism exists, in order to individually assess the roles of superior local processing and poor global integration, careful thought as to the tasks employed is required. For example, evidence of detail focus in the absence of global impairment has been observed in free drawing tasks where these two concepts are able to be independently assessed (Booth, Charlton, Hughes, & Happe, 2003).

Investigation of weak coherence as an endophenotype by exploring this trait within ASD families has made some progress. Eight studies were found where coherence tasks had been employed with siblings or parents of those with autism (see Table 3). Consistent findings on the EFT show that parents also demonstrate superior performance attending to detail, with a large effect size. The one study to employ this task with siblings found no effect (Happe, Briskman, & Frith, 2001). Results on the block design and object assembly tasks vary widely across studies, for example data from mothers on the unsegmented version of the block design ranges from a large negative effect in the direction of increased detail focus (Bolte & Poustka, 2006) to a small positive effect, in the opposite direction (Happe et al., 2001). As mentioned with regard to set-shifting (see section 3.4.1.1), the autism literature employs a wide range of comparison groups and therefore cannot be easily collated. Other methodological considerations such as differences in task administration also contribute to the heterogeneity of results.

3.4.2.2 Evidence from the eating disorder literature

Gillberg and colleagues were the first to note that a subgroup of their AN community sample (19.6%) met criteria for autistic spectrum disorder (Gillberg, Gillberg, Rastam, & Johansson, 1996), and that this subgroup showed a cognitive profile similar to those with autism. Three-year follow-up of the full cohort showed a persistent autism diagnoses in most of the previous cases (16% of full sample) in addition to persistent neuropsychological profile across all participants where a bias toward detail was seen on the object assembly task. Similar findings were presented by an Australian group, who also found a bias toward detail using the object assembly task and the group embedded figure test in 24 women with AN and 24 control women (Tokley & Kemps, 2007). Research using the Matching Familiar Figures paradigm investigated detail focus within the context of cognitive efficiency

Table 3: Summary of coherence effect sizes for 1st degree relatives of those with psychiatric illness

	ASD		OCD		BPD		MDD		SZF	
	ES	N	ES	N	ES	N	ES	N	ES	N
G/EFT	-0.73	9	-		-		-		-	
BD (segmented)	0.09	1	-		-		-		-	
BD (unsegmented)	0.04	6	-		-0.94	1	-		0.29	2
Object Assembly	0.06	2	-		-		-		-	
Fragmented Pictures	-		-		-		-		-	
ROCF organisation	-		-		-		-		-	

ASD Autistic Spectrum Disorder; OCD Obsessive-compulsive Disorder; BPD Bipolar Disorder; MDD Major Depressive Disorder; SZF Schizophrenia; ES Effect Size (unweighted mean of study effect sizes- see Appendix 2 for raw data); K n of studies contributing to ES; G/EFT Group/Embedded Figures Test; BD Block Design; ROCF Rey-Osterrieth Complex Figure

in 20 women with AN, 14 with BN and 26 HC (Southgate, Tchanturia, & Treasure, 2007). Women with AN were significantly faster at identifying the target figure (e.g. lion) from an array of eight closely matched figures, again suggesting a superiority with detailed or local processing.

Following these experimental findings, Lopez (2008) was the first to employ a hypothesis driven approach, applying the weak central coherence theory from the autism literature in the ED population. A battery of tasks were selected from those previously employed in the autism population to assess visual global processing (ROCF copy order & style indices), visual local processing (embedded figure test, unsegmented block design), and verbal processing (sentence completion test, homograph reading test). Participants were 42 women with AN, 42 women with BN, 42 women recovered from ED, and 42 healthy control women. Both AN and BN samples showed strengths on tasks requiring detailed processing and a relative weakness on tasks requiring global integration, providing evidence for criterion 1 of an endophenotype (Lopez et al., 2008b; Lopez, Tchanturia, Stahl, & Treasure, 2008d). Moreover, women recovered from an ED showed the same information processing style as those with current ED, indicating that weak coherence is not state-dependent and providing evidence for criterion 2 of an endophenotype (Lopez et al., 2008e).

A recent systematic review outlined evidence to date for the presence of weak coherence in the ED population, by synthesising the results from the literature where tasks benefiting from detail focus or global integration were employed (Lopez, Tchanturia, Stahl, & Treasure, 2008c). A number of tasks assessing coherence (Block design, Rey-Osterrieth Complex figure, Embedded figure test, Object assembly, Sentence completion test) report medium to large effect sizes in this clinical group, suggesting pronounced differences compared to healthy controls.

4   A meta-analytic review of set-shifting ability in eating disorders. [2]

4.1   Background

As described above (see 3.4.1.2), difficulties with set-shifting or cognitive flexibility has been found in those with an eating disorder (ED). The aim of this chapter was to collate and summarise the literature on set-shifting ability in women with ED, to ascertain whether difficulties shifting set is a consistent finding in the literature. A secondary aim was to identify the most sensitive measures of set-shifting in the ED population, so that these tasks could be employed in this thesis.

4.2   Method

The "QUOROM statement" for meta-analyses was followed.

4.2.1 Searching

Papers were located using the electronic databases PsycInfo, Medline and Web of Science, by additional hand searches through reference lists and specialist ED journals, and through direct contact with academic institutions with an interest in this area. Journals were searched up to December 2005. Search keyword terms were; NEUROPSYCHOLOGY, SET-SHIFTING, FLEXIBILITY, RIGIDITY, MENTAL FLEXIBILITY, COGNITIVE RIGIDITY, PERSEVERATION, WISCONSIN CARD SORTING TEST, TRAIL MAKING TEST, BRIXTON, HAPTIC, CATBAT, EATING DISORDER, ANOREXIA NERVOSA, and BULIMIA NERVOSA. No date restrictions were applied to the selection of literature.

Any study employing the set-shifting tasks Trail Making Test (TMT), Wisconsin Card Sorting Test (WCST), Brixton task, Haptic Illusion, CatBat task, or the set shifting subset of the Cambridge Neuropsychological Test Automated Battery (CANTAB) with an ED population was eligible for inclusion. All selected tasks require shifting between mental sets and strategies, although the specific operations involved may differ (a more thorough description of tasks included in this thesis is presented in chapter 5.1.1):

---

[2] Please note that a version of this chapter has been published:
Roberts, M. E., Tchanturia, K., Stahl, D., Southgate, L., & Treasure, J. (2007). A systematic review and meta-analysis of set shifting ability in eating disorders. *Psychological Medicine, 37*(8), 1075-1081.

The Trail Making task (Kravariti, Morris, Rabe-Hesketh, Murray, & Frangou, 2003). Participants connect a sequence of dots over three trials; control, alphabetical, alphanumeric switching i.e. 1 –A -2 –B -3 –C etc (Trail B). Time taken to complete trail B (switching task) is the measure of set-shifting ability.

Wisconsin Card Sort Test (WCST; Computer version 4 Psychological Corporation) Participants are instructed to match stimulus cards with one of four category cards. The sorting rule changes unpredictably during the course of the task. The number of perseverative errors is used as a measure of set-shifting ability.

The Brixton test (Burgess & Shallice, 1997). Participants are asked to predict the movements of a blue circle, which changes across 10 locations after each response (for 55 trials). The participant must adjust their responses as the pattern changes. The total number of errors is a measure of set-shifting ability.

The Haptic Illusion Task (Uznadze, 1966; Tchanturia, Serpell, Troop, & Treasure, 2001) Participants judge the relative size of two wooden balls with their eyes closed. After 15 trials of different sized balls, identically sized balls are presented. The number of illusions experienced (same sizes balls perceived as different sizes) is a measure of perceptual inflexibility.

The CatBat Task (Eliava, 1964; Tchanturia, Morris, Surguladze, & Treasure, 2002). Participants fill in missing letters in a written short story. In the first part of the story the context requires a 'C' (for CAT), then the context changes and 'B' (for BAT) becomes most appropriate. The number of perseverative errors ('C' where 'B' is appropriate) is the measure of set-shifting ability.

CANTAB IDED set shifting subtest (Downes, Roberts, Sahakian, Evenden, Morris, & Robbins, 1989). The Cambridge intra-extra dimensional (IDED) set shift consists of stimuli (colour-filled shapes and white lines) that appear in four rectangles on a computer screen. The subject must learn the correct stimuli for selection, based on audio and visual feedback. After six correct trials (maximum 50 trials) subjects move to the next stage and the rule shifts. Total number of errors is used as the measure of set shifting ability.

4.2.2   Selection

A total of 22 studies were selected following the above search criteria. Upon inspection of the full manuscripts, three of these papers were excluded (Fox, 1981; Ferraro, Wonderlich, & Jocie, 1997; Bayless et al., 2002), as raw data (mean and standard deviation) was not presented and was unavailable from the authors on

request. A further four papers were excluded as they did not contain a healthy control (HC) group, and therefore the effect size could not be calculated (Touyz, Beumont, & Johnstone, 1986; Lauer, Gorzewski, Gerlinghoff, Backmund, & Zihi, 2002; Kitabayashi et al., 2004; Frieling et al., 2005). A total of 15 papers were included in the systematic review. One of the selected papers was in a foreign language journal (Koba, Shrie, & Nabeta, 2002), and another was initially in review (Steinglass, Walsh, & Stern, 2006) but has since been published.

4.2.3 Data abstraction

Descriptive statistics (mean, standard deviation and sample size) for ED and control groups were extracted from the papers. If this data was missing it was requested from the author.

4.2.4 Quantitative data synthesis [3]

Analyses were carried out in Stata 9.1 (StataCorp, College Station, TX, USA) using the user-contributed commands for meta-analyses metan (Bradburn, Deeks, & Altman, 1988) and metabias (Steichen, 1998).

The mean difference in scores between ED and HC groups was standardised by calculating Cohen's d, the difference between the two raw means divided by the pooled standard deviation (Rosenberg, Adams, & Gurevitch, 2000). The standard error of each study's standardised effect size was calculated from the estimated effect and the group sizes of the two groups using the method by Cooper and Hedges (1994), which is implemented in metan.

Cohen's d effect sizes are defined as negligible ($\geq$ -0.15 and $<$ 0.15), small ($\geq$ 0.15 and $<$ 0.40), medium ($\geq$ 0.40 and $<$ 0.75), large ($\geq$ 0.75 and $<$ 1.10), very large, ($\geq$ 1.10 and $<$ 1.45) and huge ($\geq$ 1.45).

A meta-analysis was conducted for the TMT, WCST, CatBat and Haptic tasks (comparing ED and HC groups). The four meta analyses were conducted in the following way: The standardised effects of set shifting ability for each task was pooled using a random effects model, which assumes in addition to within group variability that the mean effects differ across studies (between study heterogeneity). Random effects models produce wider confidence intervals and are more conservative than fixed effects models but are regarded to be more realistic due to the variety of case mix and settings (Everitt, 2003). The assumption of homogeneity

---

[3] This section including writing and conducting the meta-analysis was done with the help of Dr. Daniel Stahl, consultant statistician.

of true effect sizes was assessed formally using Cochran's Q test of homogeneity. However, this test is not very powerful with small sample sizes and as a sample size independent measure of inconsistency I^2 was calculated ($I^2 = (Q\text{-df})/Q$), (Higgins, Thompson, Deeks, & Altman, 2003).

Research with statistically significant results is potentially more likely to be submitted and published than studies with non-significant results. The presence of such a publication bias for the study was assessed informally by visual inspection of funnel plots (a plot of a study's precision (1/standard error) against effect size) and formally by its statistical analogue, Begg's adjusted rank test (Begg & Mazumdar, 1994), and Egger's test (Egger, Smith, Schneider, & Minder, 1997), which are implemented in metabias.

Because of a small sample size, an average standardised effect size weighted by the inverse of the variance is presented for the Brixton task.

4.2.5   Study characteristics

All studies employed an experimental cross-sectional design. All samples included an anorexia nervosa (AN) and HC population, with four studies also including bulimia nervosa (BN) patients (Pendleton-Jones, Duncan, Brouwers, & Mirsky, 1991; Tchanturia et al., 2001; Murphy, Nutzinger, Paul, & Leplow, 2002; Tchanturia et al., 2004a), and three also including recovered AN or weight restored patients (Pendleton-Jones et al., 1991; Tchanturia et al., 2002; Tchanturia et al., 2004b). Additionally, Murphy et al. (2002) included OCD patients, and Holliday et al. (2005) included a healthy sister comparison, however these results will not be explored in this chapter. Little information on diagnosed comorbidity was given, however a number of studies reported histories of diagnosed substance abuse (e.g. Jones et al., 1991) and depression (Pendleton-Jones et al., 1991; Thompson, 1993; Ohrmann et al., 2004; Fowler, Blackwell, Jaffa, Palmer, & Robbins, 2005). See Table 4 and Table 5 for further information regarding comorbidity.

The case mix studied showed wide variation: Age, BMI, diagnosis and duration of illness were noted for each sample, in order to assess clinical heterogeneity. "Recovered AN" were classified as those who had maintained a stable BMI of 19-24 for a minimum of one year (Tchanturia et al., 2002; 2004b). "Weight restored AN" were classified as those who had maintained weight for a minimum of 6 months ($M = 47.1$ mo, $SD = 31.6$; Jones et al., 1991). "Broad AN" were classified as those groups where not all participants fulfilled the criteria for AN on weight. It

was not possible to note how many cases of AN had previous episodes of BN, or vice-versa, with the exception of the Koba paper where 50% of participants also had BN symptoms (Koba et al., 2002). Results from these subpopulations were kept separate in the analysis. See Table 4 and Table 5 for age, BMI and duration of illness details for each sample.

4.3   Results

4.3.1   Trail Making Test (TMT)

TMT was the most commonly employed measure (Witt, Ryan, & George Hsu, 1985; Pendleton-Jones et al., 1991; Thompson, 1993; Kingston, Szmukler, Andrews, Tress, & Desmond, 1996; Mathias & Kent, 1998; Murphy et al., 2002; Tchanturia et al., 2004a; Tchanturia et al., 2004b; Holliday et al., 2005; Steinglass et al., 2006). A meta-analysis of TMT shifting performance revealed a small pooled standardised mean difference of 0.36 (see Figure 3). There was no evidence of heterogeneity ($\chi^2(13) = 16.68$, $p = 0.21$) between the studies i.e. between AN and BN and with different states of severity and recovery. The effect sizes across studies were found to be consistent ($I^2 = 0.11$). Begg's funnel plot suggests that little publication bias was present (see Figure 4) and both Begg's and Eggers tests for publication bias were non significant ($p = 0.91$; $p = 0.97$, respectively). Analysis for correction of publication bias (trim and fill method) revealed little difference in the results, therefore uncorrected data is presented here.

4.3.2   Wisconsin Card Sort Test (WCST)

Five papers (Thompson, 1993; Fassino, Piero, Daga, Leombruni, Mortara, & Rovera, 2002; Koba et al., 2002; Ohrmann et al., 2004; Steinglass et al., 2006) employed the WCST with an AN population. The meta-analysis of WCST perseverative errors produced a pooled standardised mean difference of 0.62 (see Figure 5). There was no evidence for heterogeneity ($\chi^2(4) = 3.73$, $p = 0.44$) (including one study with broad criteria) or publication bias (Begg's $p = 0.81$; Eggers $p = 0.64$), and effect size was consistent across studies ($I^2 = -0.07$).

The only paper employing the WCST with a BN population was not included in the meta-analysis as raw data was unavailable (Ferraro et al., 1997) however the authors noted a significant deficit in BN compared to HC performance on this task. The BN group also displayed significantly more variance in their scores than controls.

Table 4: Demographic and effect size comparison of set shifting tasks: anorexia nervosa compared to healthy control groups

		N	Age (SD)	BMI (SD)	Comorbid	DOI (yrs)	TMT	WCST	Brixton	CatBat	Haptic	CANTAB
Steinglass et al. (2006)	AN	15	25.6 (6.0)	19.0 (1.0)	− Anx	10.8 (5.8)	0.38	0.88	-	-	-	-
	HC	11	24.0 (3.1)	22.1 (1.8)	− Dep							
Holliday et al. (2005)	AN	47	26.3 (10.2)	17.9 (2.7)	− Anx − Dep	6.0 (3.0)	0.07	-	-0.29	0.49	0.92	-
	HC	47	26.5 (6.1)	21.1 (2.3)	− OCD							
Fowler et al. (2005)	AN	25	16.9 (2.0)	15.3 (1.3)	56% Dep	2.1 (1.4)	-	-	-	-	-	0.17
	HC	25	17.6 (2.2)	22.4 (1.3)	− Anx							
Tchanturia et al. (2004b)	AN	34	27.2 (8.3)	13.7 (1.4)	− Anx − Dep − OCD	-	0.85	-	0.77	0.64	0.74	-
	HC	36	24.9 (4.8)	21.8 (1.7)								
Ohrmann et al. (2004)	AN	11	22.7 (3.8)	15.2 (1.2)	36% Dep	5.5 (5.2)	-	0.04	-	-	-	-
	HC	11	27.5 (6.3)	21.3 (2.8)								
Fassino et al. (2002)	AN	20	23.2 (6.6)	15.6 (2.2)	NR	-	-	0.62	-	-	-	-
	HC	20	23.1 (2.9)	20.6 (1.7)								
Koba et al. (2002)	AN	11	23.6 (5.7)	NR	NR	-	-	1.25	-	-	-	-
	HC	7	25.9 (4.7)	NR								
Murphy et al. (2002)	AN	16	22.3 (4.4)	14.8 (1.2)	− OCD	-	0.01	-	-	-	-	-
	HC	16	25.3 (2.6)	22.0 (2.6)								
Tchanturia et al. (2002)	AN	30	25.0 (6.7)	14.6 (2.1)	− Anx − Dep − OCD	-	-	-	-	0.57	1.15	-
	HC	23	27.6 (6.4)	21.3 (1.8)								
Tchanturia et al. (2001)	AN	15	28.1 (7.3)	14.1 (2.2)	− Anx	-	-	-	-	-	1.63	-
	HC	28	28.2 (5.6)	22.3 (2.1)	− Dep							
Mathias & Kent (1998)	AN	34	22.0 (7.4)	15.3 (1.7)	− Anx	3.4 (6.4)	0.44	-	-	-	-	-
	HC	31	20.8 (3.6)	22.8 (2.4)	− Dep							
Kingston et al. (1996)	AN	46	22.1 (6.7)	14.7 (1.7)	NR	1.7 (3.6)	0.46	-	-	-	-	-
	HC	41	22.0 (5.8)	22.1 (1.9)								

Study	Group	N			Comorbid.	DOI						
Thompson (1993)	AN	10	25.8	NR	40% Dep	-	0.93	0.50	-	-	-	-
	HC	10	23.2	NR	– OCD							
Jones et al. (1991)	AN	30	24.4 (5.3)	59.4 (6.6)	47% Dep	2.32 (0.8)	0.28	-	-	-	-	-
	HC	39	24.9 (4.4)	98.2 (7.5)	7% SA							
Witt et al. (1985)	AN	16	16.4 (1.9)	NR	NR	-	0.59	-	-	-	-	-
	HC	16	16.2 (2.0)	NR								

BMI Body Mass Index; SD Standard Deviation; Comorbid. Comorbidities; DOI Duration of Illness in years; TMT Trail Making Task B; WCST Wisconsin Card Sorting Test perseverative errors; CatBat CatBat task shift/BAT time; CANTAB Cambridge Intra-extra dimensional shift task; AN Anorexia Nervosa; HC Healthy Control; NR Not Reported; Anx Anxiety; Dep Depression; OCD Obsessive-compulsive disorder; SA Substance Abuse

Table 5: Demographic and effect size comparison of set shifting tasks: recovered/weight restored anorexia nervosa competed to healthy control groups, and bulimia nervosa compared to healthy control groups

		N	Age	BMI/IBW	Comorbid	TMT	WCST	Brixton	CatBat	Haptic	CANTAB
Recovered/weight restored AN											
Tchanturia et al. (2004b)	ANrec	18	28.4 (6.8)	20.4 (1.5)	Nill	0.45	-	0.34	0.31	0.91	-
	HC	36	24.9 (4.8)	21.8 (1.7)							
Tchanturia et al. (2002)	ANrec	16	30.0 (6.0)	20.1 (1.7)	Nill	-	-	-	-0.38	1.44	-
	HC	23	27.6 (6.4)	21.3 (1.8)							
Jones et al. (1991)	ANrec	20	26.0 (6.2)	87.8 (11.2)	NR	0.20	-	-	-	-	-
	HC	39	24.9 (4.4)	98.2 (7.5)							
BN											
Tchanturia et al. (2004a)	BN	19	26.5 (5.7)	21.8 (2.1)	– Anx – Dep – OCD	0.52	-	0.07	0.94	0.88	-
	HC	35	24.8 (4.7)	21.8 (1.7)							
Murphy et al. (2002)	BN	16	22.0 (4.5)	20.1 (2.3)	– OCD	-0.58	-	-	-	-	-
	HC	16	25.3 (2.6)	22.0 (2.6)							
Tchanturia et al. (2001)	BN	15	25.1 (7.1)	20.0 (2.3)	– Anx – Dep	-	-	-	-	1.40	-
	HC	28	28.2 (5.6)	22.3 (2.1)							
Jones et al. (1991)	BN	38	24.1 (4.0)	94.0 (7.3)	NR	0.46	-	-	-	-	-
	HC	39	24.9 (4.4)	98.2 (7.5)							

BMI Body Mass Index; IBW Ideal Body Weight %; Comorbid. Comorbidities; DOI Duration of Illness in years; TMT Trail Making Task B; WCST Wisconsin Card Sorting Test perseverative errors; CatBat CatBat task shift/BAT time; CANTAB Cambridge Intra-extra dimensional shift task; AN Anorexia Nervosa; ANrec Recovered Anorexia Nervosa; HC Healthy Control; NR Not Reported; Anx Anxiety; Dep Depression; OCD Obsessive-compulsive disorder; SA Substance Abuse.

.

Figure 3: Forrest Plot for TMT Meta-analysis

Study		Standardised Mean diff. (95% CI)	% Weight
Steinglass et al. (2006)		0.38 (-0.40,1.17)	4.0
Holliday et al. (2005)		0.07 (-0.34,0.47)	11.3
Tchanturia et al. (2004a)		0.52 (-0.05,1.08)	6.9
Tchanturia et al. (2004b)		0.85 (0.36,1.34)	8.6
Tchanturia et al. (2004b)		0.45 (-0.13,1.02)	6.8
Murphy et al. (2002)		0.01 (-0.68,0.71)	5.0
Murphy et al. (2002)		-0.58 (-1.29,0.13)	4.8
Mathias & Kent (1998)		0.44 (-0.06,0.93)	8.5
Kingston et al. (1996)		0.46 (0.04,0.89)	10.5
Thompson (1993)		0.93 (-0.00,1.85)	3.0
Jones et al. (1991)		0.28 (-0.19,0.76)	8.9
Jones et al. (1991)		0.20 (-0.34,0.74)	7.4
Jones et al. (1991)		0.46 (0.01,0.91)	9.7
Witt et al. (1985)		0.59 (-0.12,1.30)	4.8
Overall (95% CI)		0.36 (0.19,0.53)	

Standardised Mean diff.

AN
AN Recovered
AN Broad
BN

Figure 4: Begg's funnel plot (assessing publication bias) for TMT meta-analysis

Begg's funnel plot with pseudo 95% confidence limits
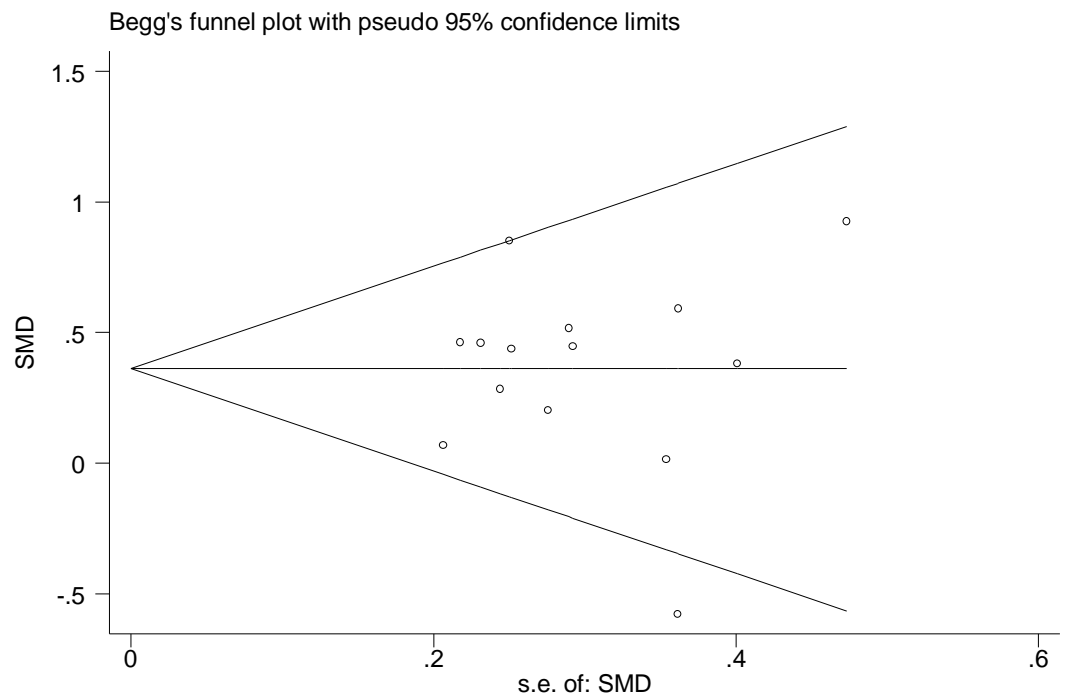
Figure 5: Forrest Plot for WCST Meta-analysis

Study	Standardised Mean diff. (95% CI)	% Weight
Steinglass et al. (2006)	0.88 (0.07,1.70)	19.7
Ohrmann et al. (2004)	0.04 (-0.80,0.88)	18.9
Fassino et al. (2002)	0.62 (-0.02,1.25)	32.7
Koba et al. (2002)	1.25 (0.21,2.29)	12.1
Thompson (1993)	0.50 (-0.39,1.39)	16.6
Overall (95% CI)	0.62 (0.25,0.98)	

Standardised Mean diff.

■ AN
▨ AN Recovered
▤ AN Broad
▨ BN
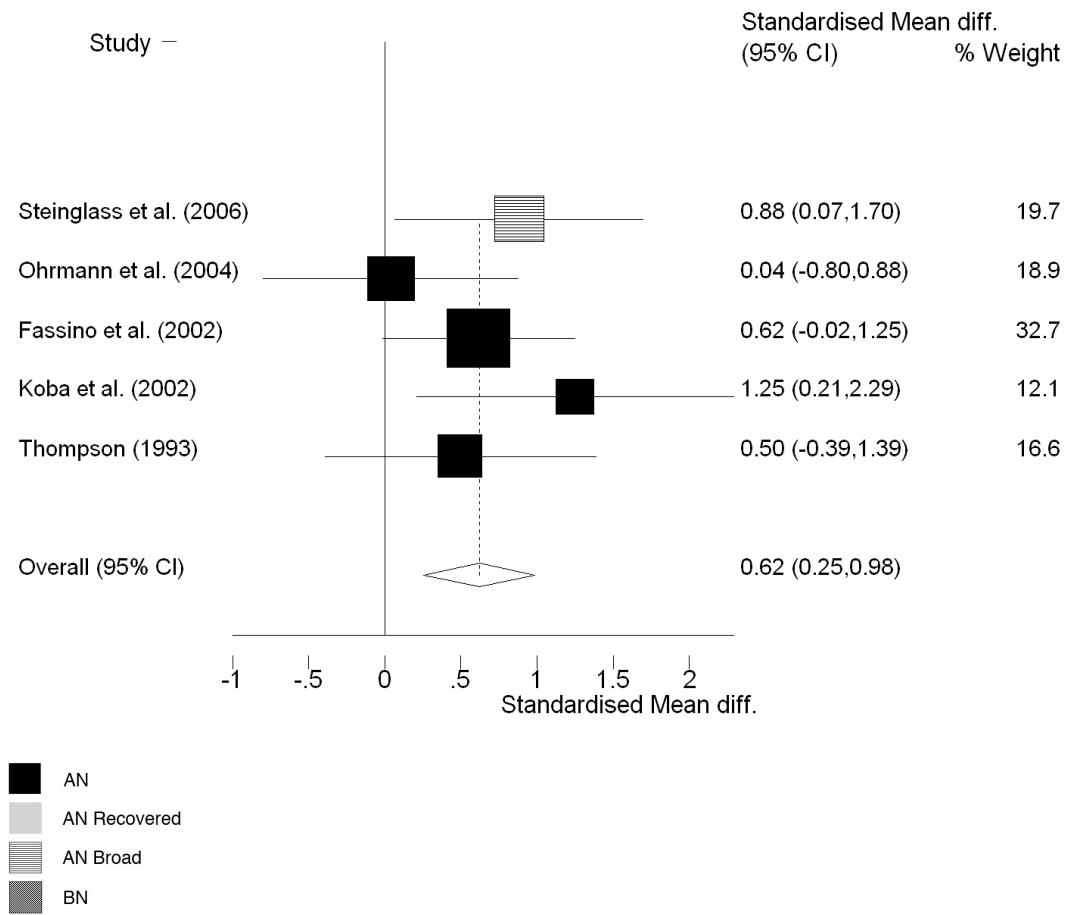
4.3.3   Brixton Task

Three published studies employed the Brixton task (Tchanturia et al., 2004a; Tchanturia et al., 2004b; Holliday et al., 2005), all from our research group. No meta-analysis was calculated for this task, as there were only four data points across studies and ED groups. An average standardised effect size of 0.21 was calculated for the Brixton task. It should be noted that wide variation in effect size was noted across samples employing this task (see Table 4 and Table 5). The only group in which the confidence interval did not overlap with zero were people acutely ill with anorexia nervosa (Tchanturia et al., 2004b).

4.3.4   Haptic Illusion

The Haptic Illusion is another measure that has only been used by our research group. A meta-analysis of errors on this task yielded a large pooled standardised mean difference of 1.05 (see Figure 6). There was no evidence for heterogeneity (χ^2 (7) = 7.11, p = 0.42) between BN or AN samples, or in AN samples with broad criteria or weight recovery. Evidence of publication bias was found (Begg's p = 0.03; Eggers p = 0.01), however this is within the 95% confidence interval limits (see Figure 7), and the trim and fill method did not predict any change in the data. Also, given the large overall effect size, it can be concluded that this finding is reliable. Effect size was consistent across studies (I^2 = 0.016).

4.3.5   CatBat Task

The CatBat task is the third measure that has only been used by our research group (Tchanturia et al., 2002; Tchanturia et al., 2004a; Tchanturia et al., 2004b; Holliday et al., 2005). A meta-analysis of CatBat performance revealed a medium pooled standardised mean difference of 0.45 (see Figure 8). Heterogeneity was non significant. No evidence for publication bias was found (Begg's p = 0.71; Eggers p = 0.47).

4.3.6   CANTAB set-shifting

Only one study was found that employed the CANTAB IDED shifting sub-test in an eating disorder population (Fowler et al., 2005). No difference was found between 25 AN and 25 healthy control participants, with a small effect size of 0.17.

Figure 6: Forrest Plot for Haptic Meta-analysis
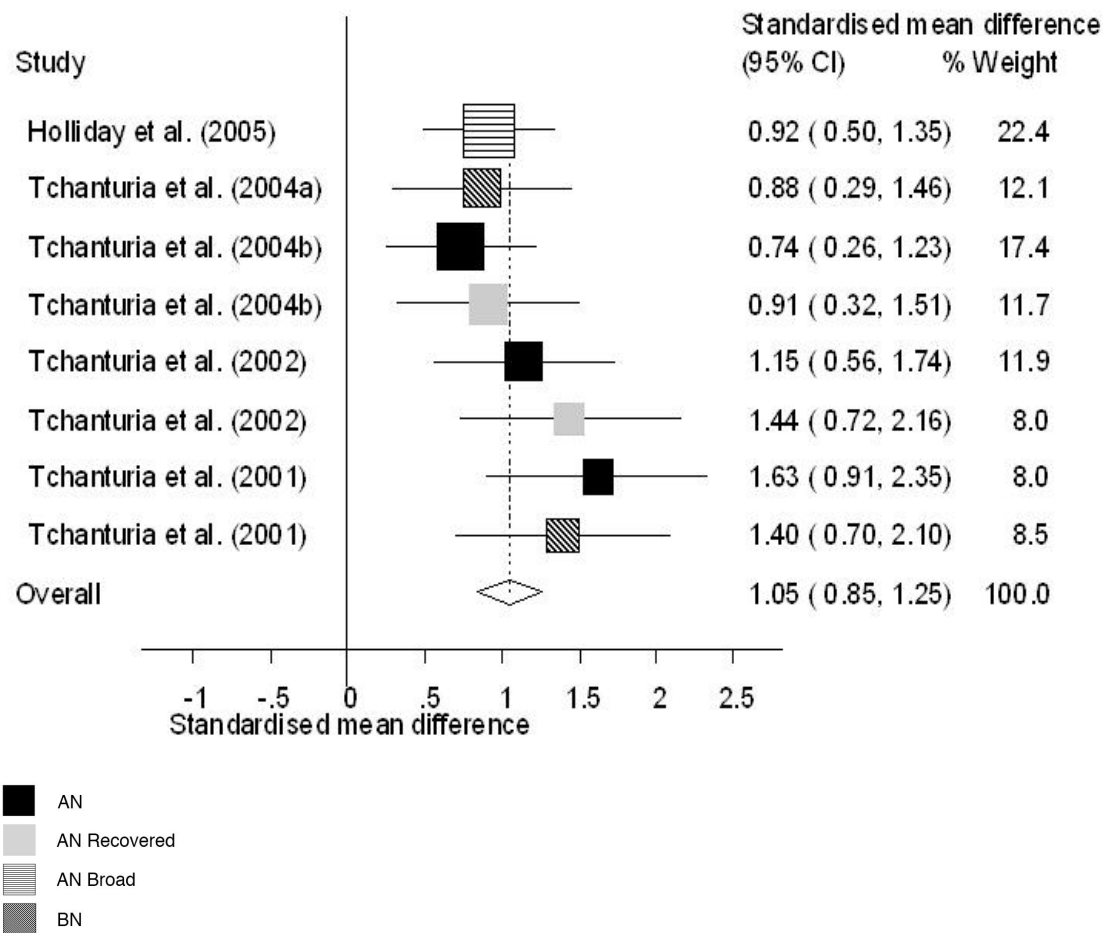
Figure 7: Begg's funnel plot (assessing publication bias) for Haptic meta-analysis

Begg's funnel plot with pseudo 95% confidence limits
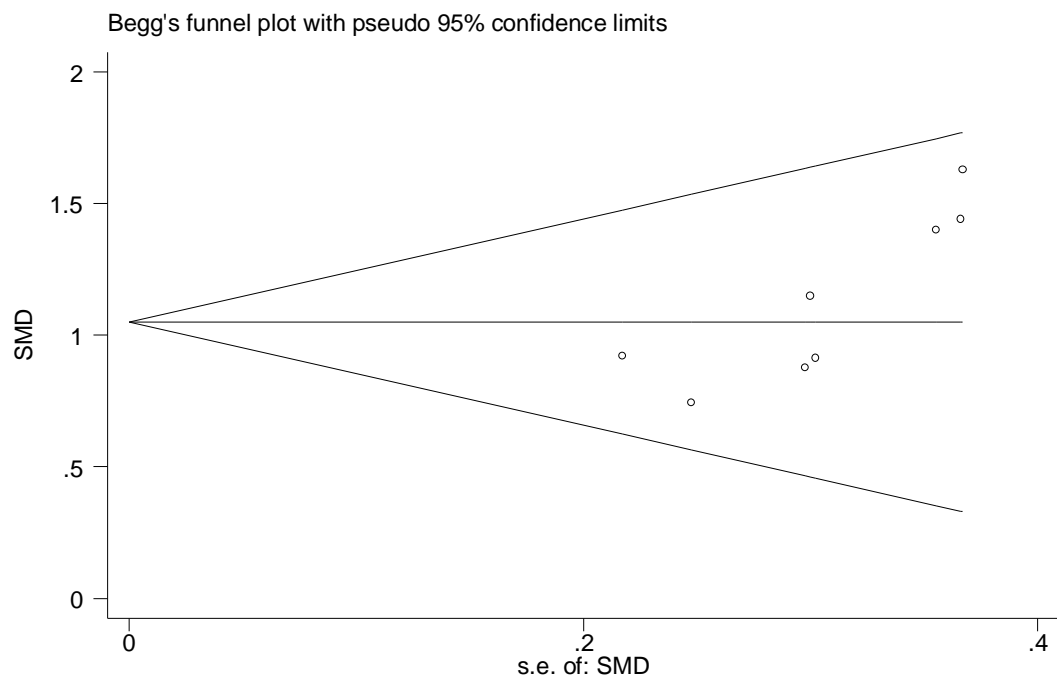
Figure 8: Forrest plot for CatBat Task meta-analysis
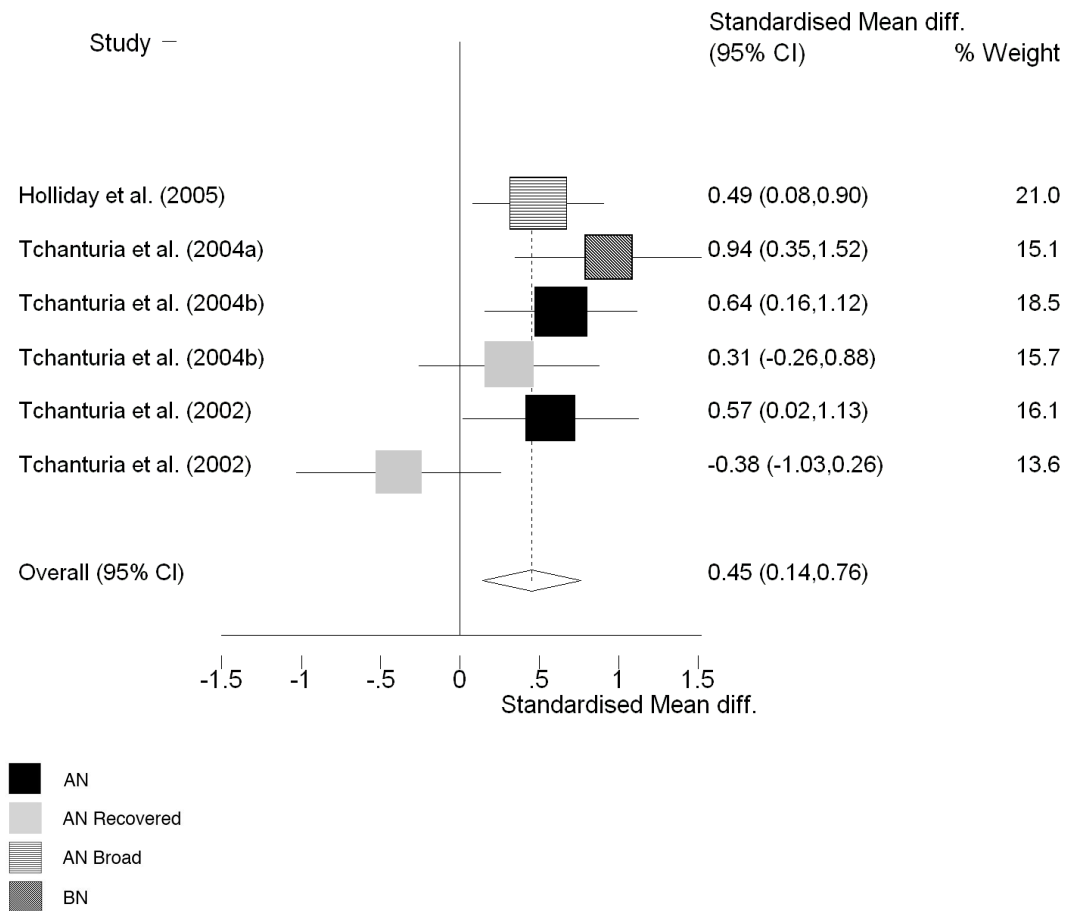
4.4   Discussion

This chapter reviewed 15 studies that administered at least one of six neuropsychological set-shifting tasks in eating disorder populations. A consistent deficit in set shifting ability was found across diagnoses, state of illness and most of the set-shifting measures. It was possible to conduct a meta-analysis of studies for TMT, WCST, Haptic, and CatBat tasks. The size of the pooled effect size varied between tasks, from small (TMT B), to medium (WCST and CatBat Task), to large (Haptic Task). The Brixton task has been less widely used and may only show an effect in the acute state. The set-shifting subtest of the CANTAB was used once and had an effect size close to negligible.

The limited amount of data from the recovered/weight restored subgroups of AN suggested that that the deficit in set-shifting in some tests (TMT, Haptic, CatBat) remains as a trait and might be a candidate endophenotype. Further research is required toinvestigate this possibility. Likewise, the available data from people with BN was restricted, but suggests that the deficit in set-shifting measured with TMT, CatBat Task, and Haptic Task is similar to that of anorexia nervosa whereas the Brixton task showed no effect.

There appears to be a difference in the potency of the set-shifting tasks that have been used. The Haptic task clearly has the highest, most consistent effect sizes. It is interesting to note that was the only task employed where set-shifting was measured perceptually. Grunwald et al. (2001a; 2001b) assessed haptic performance in AN (before and after weight recovery) and control women by asking them to reproduce a tactile pattern they had traced with their fingers. The drawings of AN women were of considerably poorer quality regardless of state of illness, suggesting a general deficit in somatosensory or haptic processing. It is possible that as these haptic tasks are initially more difficult, administering a haptic task as a measure of set-shifting served to magnify effect size, therefore producing the larger effect sizes seen for the Haptic task in this paper.

From the limited data available the deficit in set shifting is found across eating disorders diagnoses. Interestingly this deficit is also found in other psychiatric conditions. We searched for meta analysis/systematic review/review and ADHD, psychiatric, manic depressive, psychosis etc and found that a set-shifting deficit is not specific to the eating disorder population: In a systematic review of cognitive deficits including set-shifting in euthymic patients with bipolar disorder, effect sizes

were between 0.5 and 0.8 (Robinson et al., 2006), and in adult ADHD effect size was 0.65. Set shifting abnormalities have also been found in the first degree relatives of people with bipolar 1 disorder (Clark, Sarna, & Goodwin, 2005) and of schizophrenia (Snitz, Macdonald III, & Carter, 2006). Effect size for the WCST was small for people with OCD (Henry, 2006), who display a larger effect size on the CANTAB IDED shift (Chamberlain, Fineberg, Blackwell, Robbins, & Sahakian, 2006). Thus it appears that weak set-shifting is an endophenotype that broadly increases the risk of many forms of psychiatric illnesses.

A number of limitations in the current literature have been identified throughout this review. Firstly, the majority of studies employ small samples. Additionally, we excluded longitudinal data sets as there is no evidence to our knowledge that these tasks are reliable if used repeatedly. Furthermore, the Haptic, Brixton, and CatBat tasks have been employed only by our research group to date. Replication by other research groups among differing samples is required in order to validate the findings presented here. Finally, it is unfortunate that only one study employed the CANTAB IDED shifting task. An explanation for the small effect size of this task could found in the case mix of the Fowler et al. (2005) study. This population differs markedly from the other populations presented here, as it is an adolescent group with a short duration of illness. This is relevant because the diagnosis of AN is unstable in the early phase, as many cases recover or evolve into BN. These uncertainties exemplify the need for further work in this area.

4.4.1   Conclusions

This review highlights a consistent set-shifting deficit in the ED population, providing the baseline evidence for a formal investigation of set-shifting as an endophenotype in this population. As current treatment attempts for AN are largely ineffective, understanding deficits in neuropsychological functioning becomes increasingly important given the potential application of these findings to the clinical setting (Tchanturia, Davies, & Campbell, 2007). Moving the focus away from visible behaviours to concentrate on these underlying cognitive traits may provide the way forward for novel treatment approaches.

5   Methodology

This chapter will outline the general methodology across all four empirical studies presented in this thesis. Details are given for the three types of measures employed; neuropsychological assessment, clinical interviews, and self-report questionnaires. A justification of the selection of neuropsychological tasks is outlined. The study protocol and other methodological details are also presented.

5.1   Neuropsychological Assessment

In order to assist training and dissemination of the following neuropsychological battery, administration and scoring protocols were drawn up for all seven tasks (see Appendix 3). These protocols detail experimenter instructions for set-up, administration, and scoring of all set-shifting and central coherence tasks outlined below.

5.1.1   Set-shifting

5.1.1.1   Selection of set-shifting tasks for this thesis

In order to include the most sensitive and reliable set-shifting tasks for this thesis, tasks with medium or large standardised mean differences from the meta-analyses reported in chapter 4 were automatically included (WCST, Haptic, CatBat). While the TMT displayed only a small effect size, given that it is the most frequently employed measure of set-shifting ability and has been in reported in the eating disorder (ED) literature as early as 1985, it seemed necessary to include this task also, both in order to directly compare this cohort with earlier TMT results and to compare the performance of this traditional measure with more recent tasks of set-shifting. Of the remaining two tasks, the CANTAB shift was discarded due to poor evidence of sensitivity and logistical difficulties (the CANTAB is an automated neuropsychological battery for which purchasing a licence was outside the budget of this thesis). Despite a small overall effect size, large variation in performance was observed for the Brixton task across studies. It was decided to include this task due to its potential as a sensitive measure in acute AN (Tchanturia et al., 2004b), and because of its quick (approximately three minute) administration time. This gave a total of five set-shifting tasks, measuring the construct from a variety of different mediums; three purely cognitive (TMT, WCST, Brixton), one linguistic (CatBat), and one perceptual (Haptic).

5.1.1.2   Trail Making Test (TMT)

The computerised TMT (Kravariti et al., 2003) consists of three tasks; jumping flea, alphabetic sequences, and alphanumeric sequences. For each task participants are given a demonstration with 6 targets, a practise trial with 6 targets, then the real trial with 20 targets. For the first task, participants follow a jumping red circle on the computer screen by clicking on it as fast as possible in each different position. In the second task the circles are lettered (A through T), and participants must alphabetically connect circles in a "dot-to-dot" fashion i.e. A - B - C etc. For the final task circles are numbered or lettered, and they must alternatively link numbers and letters i.e. 1 - A - 2 - B - 3 – C. This third task is equivalent to "Trail B" of the pen and paper version of the TMT. Mean and total reaction times are recorded for each of the trials, along with the number of incorrect responses i.e. where D was clicked instead of C (in alphabetic sequences), or D was clicked instead of 4 (in alphanumeric sequences).

The TMT is one of the oldest neuropsychological measures in the literature, with papers describing this task over half a century ago (Reitan, 1955, 1958; Reitan & Tarshes, 1959). The original pen and paper version was first developed as part of the Army Individual Test Battery (1944), with the computerised version used here developed at the Institute of Psychiatry in 2003. The TMT has been well used in mental health populations, such as depression (Fossati, Ergis, & Allilaire, 2002) and schizophrenia (Perianez et al., 2007). It remains one of the most widely used neuropsychological measures, perhaps because it is a quick and simple measure to administer and score. Traditionally, the time taken to complete trail B (shifting trail) has been used as the measure of set-shifting ability. This outcome is presented here however with the introduction of a computerised version of the TMT for increased reaction time accuracy, an additional variable has come into play. While it is unusual in modern times to be unfamiliar with a computer, an individual's competency when using a mouse is a real factor in this task, given that the main outcome is a measure of reaction time which is naturally influenced by the speed with which the participant can control a mouse. Thus alongside the traditional reaction time outcome for trail B (shifting), this thesis also presents a TMT ratio, where the total time for the shifting trial is divided by the non-shifting trial (alphabetical sort), in order use an individuals baseline time to control for variations in mouse speed. This should produce a more

refined measure of the time deficit incurred when a shifting component is added to the task.

5.1.1.3 Wisconsin Card Sorting Test (WCST)

Subjects are instructed to match stimulus cards presented in a single pile at the bottom of a computer screen with one of four category cards (or key cards) presented in a row at the top of the screen; a single red triangle, two green stars, three yellow crosses and four blue circles. These four category cards remain at the top of the screen throughout the task, with the cards sorted by the participant stacking in corresponding piles underneath the category cards. Each card can be sorted by one of three rules; colour, shape, or number. Participants must match each stimulus card they are given by using a mouse to click on the category card representing the sort rule they have chosen (e.g. clicking on the single red triangle to sort in to the 'red' pile). Once a pile is selected, each card takes 1.5 seconds to sort, during which the participant can change their mind by clicking once again anywhere on the screen. However once a card has reached its destination pile, it cannot be moved. After each card sort, feedback is given in the form of the word 'RIGHT' or 'WRONG' flashing on the computer screen for 1 second. At the same time as the visual feedback, a male voice announces whether the sort was right or wrong. Participants are to use this audio/visual feedback to help them sort the next card correctly, with the aim to get as many correct responses as possible. To start with, the designated sort rule is colour, however this rule changes seemingly unpredictably during the course of the task (colour, shape, number, colour, shape, number). Once the participant has completed 10 correct sorts in a row, the sorting rule changes in the order detailed above. Administration is terminated when either a) the participant correctly sorts all six category rotations, or b) the maximum number of cards (128) are administered. It is possible to complete the task with only 70 cards. There is no time restriction on the task.

The WCST was originally developed in 1984 as a manually administered task, presented on a table top using cardboard key cards (Grant & Berg, 1984). Participants would sort cards by hand into the four piles outlined above, with the experimenter coding each sort and giving verbal feedback (right/wrong). In 1993 the WCST was standardised by Heaton and colleagues, who created the computerised version (Heaton, Chelune, Talley, Kay, & Curtiss, 1993).

While the WCST is generally regarded as a set-shifting measure, it does in fact measure multiple different aspects of executive functioning including working memory, reward, and inhibition. Thus rather than being a specific shifting task, it is more accurately a general measure of frontal functioning. However as the WCST produces a number of different outcome measures (Trials administered; Total correct; Total errors; Perseverative responses; Perseverative errors; Non-perseverative responses; Categories completed; Trials to complete 1st category; Failure to maintain set; Learning to learn), it is possible to extract the variables of interest in order to have a more pure assessment of shifting ability. Outcome measures specific to set-shifting used in this study are firstly number of perseverative errors (when a sort is made based on a previous sorting rule that is no longer relevant), and secondly number of categories completed (where completing a category is 10 correct sorts e.g. 10 correct colour sorts, after which the sorting rule changes).

Arguably the most sensitive outcome measure of set-shifting from the WCST is that of perseverative errors. The WCST presents perseverative errors as both a raw score (n of errors) and as a percentage score of the total number of cards administered. In this thesis the raw score is used, as the more cards a participant is administered, the more masked their perseverative responses become when using this percentage score. This is because the more errors (perseverative or otherwise) the participant makes, the more cards they are administered in order to complete the task. Each individual error therefore counts as an increasingly smaller percentage as more cards are administered. For example, a participant who completed the WCST in 80 cards with 10 perseverative errors gets a 12.5% score. However a participant making the same number of perseverative errors but needing 128 cards (due to other types of errors) gets a 7.8% score. In this second case, random errors effect the participants perseverative error percentage score by inflating the number of trials administered and thus decreasing the contribution of each perseverative error to the percentage score. Therefore, in order to use the purest measure of set-shifting for the WCST, the number of raw perseverative errors are used in this thesis.

5.1.1.4 Brixton Task

Participants are presented with a computer screen displaying two horizontal rows consisting of 5 circles each. Each circle is numbered directly underneath, numerically from right to left on the top row (1-5), then right to left on the bottom

row (6-10). The first circle (numbered 1) is coloured blue, while the rest of the circles are blank/white. Participants are told that the blue circle will change location from circles 1 through 10 on the experimenters mouse click. Participants are asked to predict where the blue circle will appear next, by saying the number of the prediction out loud (e.g. "circle 4"). After the first trial a pattern in the movement of the dot emerges. Participants are told that occasionally, this pattern of movement changes without warning and they must abandon the old concept in favour of a new one (e.g. 2-3-4-5-6-*5-4-3-2-1*). The pattern changes eight times over the 55 trials, and the number of errors made is recorded (shifting outcome measure). On each pattern change, the first trial of the new pattern is marked correct if the participant's prediction is what would be expected from the previous rule (e.g. following the above circle movement, the correct responses would be 2-3-4-5-6-*7-4-3-2-1*). By taking this first trial into account, the resulting incorrect responses are all indicative of shifting errors.

The Brixton task was developed by Burgess and Shallice (Burgess & Shallice, 1997) as part of the Brixton & Hayling task set. The original test was a paper version, where A4 sheets were turned for each new trial (in place of the mouse click).

5.1.1.5   CatBat Task

Participants are presented with a short story on a piece of A4 paper. Some of the words in the story have the first letter missing (e.g. _at). Participants are asked to fill in missing letters, so that the story makes sense. In the first part of the story the context requires a 'C' (for Cat), then in the second half the context changes and 'B' (for Bat) becomes most appropriate. The number of perseverative errors ('C' where 'B' is appropriate) and the time taken to complete each block of the story ('Cat' time, and 'Bat' time) is recorded. Participants are discouraged from reading ahead in the story. A copy of the task can be found in Appendix 4. The CatBat produces 5 outcome variables; CAT time (time in seconds from first Cat completion to last Cat completion), BAT time (time from last Cat completion to last Bat completion), total time (time from first Cat completion to last Bat completion), Ratio (BAT/CAT time), and number of perseverative errors (number of errors in the BAT section of the story, i.e. where a Cat response was made). The perseverative error tally includes only the participants first response to each completion in the BAT section. Therefore whether

the participant changes an initial Cat completion to Bat, or leaves the incorrect Cat response in the story, an error is counted.

The CatBat task was adapted by our team at the eating disorders unit from a similar task developed in Georgia (Eliava, 1964). It has been part of our assessment battery for over 5 years, with consistent results from the task being presented in five publications. American collaborators have recently adopted the task.

5.1.1.6 Haptic Illusion

This perceptual set-shifting task (Uznadze, 1966) uses three wooden balls of equal weight but different sizes (two balls of the same size, and one larger ball). Participants are asked to close their eyes, place their hands palm up on their knees, and asked to give a judgement about the comparative size of the two wooden balls to be placed in either hand (i.e. are they approximately the same size, or is one bigger or smaller). Balls are rolled into the participant's hands from the fingertip to their palm (encouraging the participant to wrap their fingers up around each ball) then immediately rolled out again. Participants are first given same-sized balls in each hand, to test for asymmetry. Secondly, 15 trials with balls of different sizes are administered (fixation stage), where the participant gives a size judgement but this is not recorded. By default the bigger ball is placed in the participants left hand, however if in the asymmetry stage the participant reports feeling the ball in their right hand to be larger, the bigger ball is placed in the right hand to counteract their asymmetry. In the third stage the larger ball is discarded and participants again judge the relative size of the two same sized balls (critical stage). The number of trials where illusions are experienced (i.e. the same sized balls are perceived as different sizes) is recorded, up to a maximum of 30 trials.

5.1.2 Central Coherence

5.1.2.1 Selection of coherence tasks for this thesis

From the systematic review of central coherence in ED (Lopez et al., 2008c), four tasks emerged as most sensitive in the ED population (2 for each detail focus and global integration). The block design task and EFT showed the highest effect sizes, where a clear illustration of the effect of removing the memory component of the EFT is seen. When the EFT includes a memory component, working memory becomes a large factor in this task and performance is substantially impaired compared to controls. However when the memory component is removed (the target

shape remains visible throughout each trial), a clearer picture of detail focus is seen, where performance is significantly superior to that of controls. Given the larger effect size of the EFT and following consultation with colleagues previously employing these tasks in the ED population, the EFT (without memory component) was chosen as the measure of detail focus for the current thesis.

The Rey-Osterreith complex figure (organisation strategy; ROCF) and the sentence completion task (SCT) emerged as the most sensitive measures benefiting from global integration. The SCT is a robust measure of global deficit in the ASD population, where a large number of local completions are seen for example "he work up when it was day and NIGHT" compared to a correct (global) completion "he woke up when it was day and WENT TO SCHOOL". However discussion with colleagues revealed that local completions are rarely seen in the ED population where the outcome measure reported is reaction time rather than number of local completions. Here it is assumed that a longer reaction time indicates a local completion was thought of, but then corrected before giving a global answer. This assumption makes the SCT problematic, as a number of other factors of executive function such as inhibition and set-shifting become a real factor, clouding the intended variable of global integration. Therefore the ROCF, a measure relatively free from task constraints, was chosen as a measure of global integration. It should be noted that the ROCF is a classic case of both global integration and detail focus being simultaneously tapped. However no other more pure task of global integration was found that had been trialled in the ED population, therefore the ROCF and the EFT are selected for the current thesis.

5.1.2.2   Rey-Osterrieth Complex Figure (ROCF)

This complex figure task is used as means of examining the organisational style with regard to copying of a complex figure. Participants are provided with a blank sheet of paper, and 10 colour pencils sorted into the following order; black, green, purple, brown, blue, pink, light blue, red, orange, yellow. Starting with the black pencil, they are asked to copy the Rey-Osterrieth figure as carefully as they can. The figure remains visible to the participant at all times. The experimenter prompts the participant to change to the next colour pencil as the copy trial progresses, by placing the new colour pencil in front of the participant once they have finished one or two elements of the drawing. Additionally, the experimenter videos the drawing for scoring purposes. The delayed recall trial is administered 20

minutes after the participant had finished the copy trial. The Taylor system (Spreen & Strauss, 1998) is used to give an accuracy score between 0-36 for each drawing, where a higher score indicates a more accurately constructed drawing. Additionally, a scoring system developed by Booth and colleagues is used to measure central coherence (Booth, 2006). The latter scoring system produces a 'central coherence index' (0-2) resulting from independent scores for order and style of construction, where higher scores indicate a more coherent drawing style.

ROCF Accuracy scoring- A score between 0 and 36 is allocated to each drawing. This score is derived by dividing the drawing into 18 distinct elements (detailed in Figure 9) and giving each element a score of 2 (element correctly drawn & correctly placed), 1 (element correctly drawn but incorrectly placed, or incorrectly drawn but correctly placed), 0.5 (element incorrectly drawn & incorrectly placed, but recognisable) or 0 (element not present). The Taylor system provides a description of how each element should present in the figure, which forms the basis of rating the accuracy score of each element. For example element 4 (horizontal line within the large rectangle) *"must clearly go from the midpoint of the left hand side of the rectangle to the midpoint of the right hand side of the rectangle in one straight line"* (for a full list of element descriptions, see Appendix 3). Scores for each of the 18 elements are totalled to give the full accuracy score. A score of 36 indicates a perfectly constructed drawing.

ROCF central coherence scoring – The central coherence index is a score between 0 and 2 that indicates whether participants employ a more local (detailed) or global ('big picture') strategy when drawing the ROCF. This score is derived from assessing the order and style of the drawing (detailed below), and adding these weighted scores to produce the total central coherence index. The same method is used for both copy and recall trials.

ROCF central coherence (order)– 'Order' is a quantitative representation of whether construction of the drawing began with more detailed elements, or more global elements. Each of the 18 elements are divided into four categories (see Table 6); global external element (score = 4, e.g. large rectangle), global internal element (score = 3, e.g. horizontal line), local perimeter element (score = 1, e.g. horizontal cross), and local internal element (score = 0, e.g. circle with three dots). For each drawing, the first six elements completed by the participant are identified, and

Figure 9: Elements of the Rey-Osterrieth Complex Figure
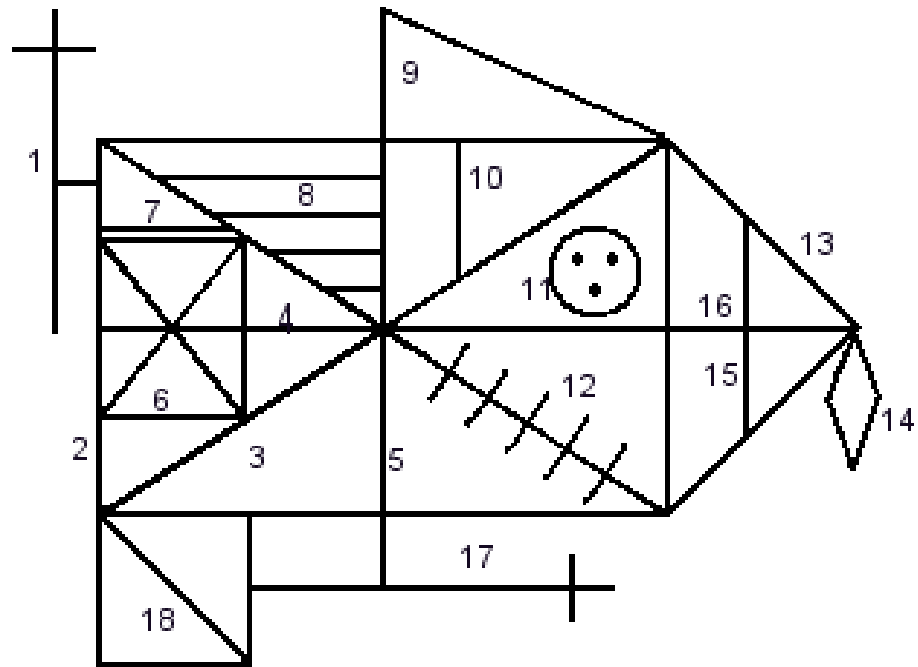
Table 6: Hierarchical categories for order of construction indices for Rey-Osterrieth Complex Figure

Category	Element	Description
Global external element (score=4)	2	Large Rectangle
	13	Sides of the large triangle attached to large rectangle
Global internal element (score=3)	3	Diagonal cross
	4	Horizontal midline of large rectangle
	5	Vertical midline of large rectangle
	16	Horizontal line within sides of large triangle
Local perimeter element (score=1)	1	Vertical cross
	9	Small triangle above large rectangle
	14	Diamond
	17	Horizontal cross
	18	Square attached to large rectangle
Local internal element (score=0)	6	Small rectangle
	7	Small horizontal line above small rectangle
	15	Vertical line within sides of large triangle
	8	Four parallel lines
	10	Small vertical line within large rectangle
	11	Circle with three dots
	12	Five parallel lines

From (Booth, 2006)

allocated a score depending on which of the above four categories those six elements fall into. The mean score across these six elements is calculated, giving a score between 0 and 3.2. This score is then divided by 3.2 to give a weighted order index between 0 and 1.

ROCF central coherence (style)– 'Style' is a quantitative representation of whether six key elements have been constructed in a continuous or a fragmented fashion. These elements are the large rectangle (element 2), the diagonal cross (element 3), the extended horizontal line (elements 4 & 16), the extended vertical line (element 5 and either the element extending above or below element 5- whichever yields the higher score), the sides of the large triangle (element 13), and the small rectangle (element 6). A score between 0 and 2 is given to each of these elements, where 2 indicates a coherent, continuously drawn element, 1 indicates a partially fragmented element (e.g. drawn with one interruption), and 0 indicates a fragmented element with 2 or more interruptions in its drawing. If an element is not present, it is not scored 0 (as this would indicate the element was present and fragmented) but is not allocated a score. Again, the mean across these six ratings is calculated, this time giving a score between 0 and 2. This score is then divided by 2, to give a weighted style index between 0 and 1. The weighted style index is added to the weighted order index, giving the total central coherence index between 0 (detail focus) and 2 (global perception).

5.1.2.3 Group Embedded Figures Task (GEFT)

The GEFT (Witkin, Oltman, Raskin, & Karp, 2002) requires participants to find and trace where a simple shape is hidden within a more complex shape. Twenty-seven complex shapes are presented two per page in a task booklet split into three sections: a practise section with nine trials, and two test sections also with nine trials each. The eight different simple shapes are presented on the back page of the booklet. For the purposes of this study, the task is modified slightly (following the provedure used by Booth, 2006) by making the simple shapes always visible to the participant (a second booklet with the back page facing up was placed beside the participant) in order to reduce the memory component of the task. Participants are first shown the target simple shape, and then the complex shape. A stopwatch is started as soon as the more complex shape is revealed. As there are two shapes per page, one is always covered by a blank piece of card; therefore the complex shape is revealed by either turning the page to reveal the top shape (bottom shape is covered

with card) or by moving the card from the bottom half to the top half of the page, thus revealing the bottom shape. When the participant thinks they have found the shape the stopwatch is paused, and the participant traces in the shape with a colour pencil. If the participant was incorrect, the stopwatch is re-started and the error is marked as a 'false claim'. Each trial is completed when a) the participant finds the target shape, or b) 60 seconds of time has passed (time-out error). Time taken to find where each of the simple shapes are hidden within the more complex shapes are recorded, along with the total number of "false claims" (when the participant says they have found the shape, but they were incorrect), and the number of time-out errors. As this task has a ceiling for response times (60 seconds) the median time of the 18 test shapes is used as the main outcome measure.

The GEFT was developed to replace the Embedded Figure Task (EFT) forms A and B (Witkin, 1971) to include the possibility of group administration. The EFT was developed from early work by Gottschaldt, who embedded shapes within line patterns in order to study perception (Gottschaldt, 1926) . This concept was built on in the development of the EFT, where two sets of 12 figures were developed. Parts of the complex figures were coloured in order to further obscure the hidden simple shape. The GEFT consists of both figures from the EFT and figures from the original Gottschaldt studies, with blue shading used in place of various colours as in the EFT. The GEFT was validated in a sample of college students (Witkin et al., 2002).

5.2   Clinical Interviews

5.2.1   SCID Extended Module H

The Structured Clinical Interview for DSM-IV diagnosis (SCID) extended module H was used to make an accurate clinical diagnosis of both current and lifetime eating disorder. This extended module was developed for the multi-centre NIMH funded Genetics of Anorexia Nervosa (GAN) study, which our unit was a part of. The interview must be conducted by a trained clinical interviewer, either over the phone or face-to-face. Administration time was between 15-60 minutes, depending on the complexity of each case. Both current and lifetime AN, BN, binge eating disorder and EDNOS were assessed. Diagnostic algorithms follow that of the DSM-IV. Specific details regarding each diagnostic category are detailed below.

5.2.1.1 Anorexia nervosa diagnosis

Participants must meet criteria A (refusal to maintain body weight at approximately 85% of that required for height), B (fear of fatness), C (at least 1 of C1, C2 or C3; disturbance in the way body weight/shape is experienced; undue influence of weight/shape on self-esteem; denial of the seriousness of low weight) and D (amenorrhea) in order to merit this diagnosis. In some cases criterion D was difficult to rate e.g. prolonged irregular menses, use of the pill, inaccurate memory. In such instances clinical judgement was used. If the participant was on the pill at the time of their lowest weight, criterion D was rated 3, or threshold. AN diagnoses were subtyped into restricting type (ANR), purging type (ANP) and binge/purging type (ANBP). ANR were those who used fasting and/or excessive exercise to control their weight. ANP were those who used any type of purging method such as self induced vomiting, laxatives, or diet pills. There was no frequency of purging required for this subtype other than any of these methods being used successfully more than once (i.e. attempting to induce vomiting without success was not counted). To merit a diagnosis of ANBP, criteria for a binge (BN criteria A & B) was met in addition to purging being present. Again there was no frequency criterion other than a binge episode occurring more than once. For the purposes of data analysis, ANP and ANBP subtypes were combined.

5.2.1.2 Bulimia nervosa diagnosis

Participants must meet criteria A (eating within a discrete period of time e.g. 2 hours, an amount of food definitely larger than most people would eat during a similar time/under the same circumstances), B (loss of control over eating), C (C1 & C2), D (self-evaluation is unduly influenced by shape/weight) and E (not during a period of AN) to merit this diagnosis. Criteria C1 (binge frequency) must be a minimum frequency of 2x per week over a 3-month period. If this criterion is not fully met, a diagnosis of EDNOS is considered. Likewise criterion C2 (inappropriate compensatory behaviour frequency) must be a minimum frequency of 2x per week over a 3-month period. If this criterion is not met, a diagnosis is of EDNOS is considered. No subtyping was made for BN diagnosis.

5.2.1.3 Binge eating disorder diagnosis

BED diagnostic criteria are the same (A-E) as those for BN however criterion C2 (inappropriate compensatory behaviour) must be rated as 1 (absent). Additionally criterion C1 (binge frequency) is adjusted in that the participant must have 6 months

(not 3 months as in BN) of binge episodes occurring at least 2x per week in order to meet this criteria.

5.2.1.4 EDNOS diagnosis

A diagnosis of eating disorder not otherwise specified (EDNOS) is given to those who meet all but one of the criteria for AN, BN or BED. The most common example in the clinical group presented here is sub-threshold BN, where the frequency criterion (C1) for binge eating in BN is not met. It should be noted that in all cases included in the current study, EDNOS was only ever given as a secondary diagnosis i.e. the participant met full criteria for AN-P and received a secondary EDNOS diagnosis for sub-threshold bulimia. In such cases because the BN was subthreshold, the participant was included in the AN-P group, not the ANBN group. The only exception to this rule is made in Study 3, where one sister in three sister pairs concordant for an eating disorder only met for EDNOS. These three sisters with EDNOS were not included in any other analysis.

5.2.2 SCID-I

The Structured Clinical Interview for DSM-IV diagnosis (SCID) is a clinician administered interview used to assess current and lifetime psychiatric diagnosis. The SCID can only be administered by trained clinical interviewers, and is considered the gold standard of diagnostic assessment. For the purposes of the current study, modules A, E and F were used (in addition to the extended module H described above) to assess comorbid mood, substance, and anxiety disorders. Participants were assessed for alcohol and drug abuse and dependence, panic disorder (with/without agoraphobia), social phobia, specific phobia, obsessive-compulsive disorder, post-traumatic stress disorder, generalised anxiety disorder, body dysmorphic disorder, major depressive disorder, bipolar disorder, dysthymia and cyclothymia. Self-harming and psychotic behaviour were also recorded in the context of this interview.

For the purposes of this study, both threshold (all criteria met) and subthreshold (all but 1 criteria met) diagnoses are given. In the case that criterion D ("[diagnosis] is not better accounted for by another mental disorder") is the one criteria not met, then no subthreshold diagnosis is given. In addition to the presence of a diagnosis, severity, age of onset, and duration of illness are also recorded.

5.2.2.1 Severity

For each threshold or subthreshold diagnosis on the SCID, participants are given a severity rating of 1-6 as follows: 1= Severe (*"Many symptoms in excess of*

those required to make the diagnosis, or several symptoms that are particularly severe, are present, or the symptoms result in marked impairment in social or occupational functioning"), 2= Moderate (*"Symptoms or functional impairment between 'mild' and 'severe' are present")*, 3= Mild (*"Few, if any symptoms in excess of those required to make the diagnosis are present, and symptoms result in no more than minor impairments in social and occupational functioning")*, 4= Partial remission (*"The full criteria for the disorder were previously met but currently only some of the symptoms or signs of the disorder remain")*, 5= Full remission (*"There are no longer any symptoms or signs of the disorder but it is still clinically relevant to note the disorder")*, 6= Prior history (*"There is a history of the criteria having been met for the disorder but the individual is considered to have recovered from it")*. It should be noted that for an individual to be classified as recovered from AN or BN, they must have a severity rating of 5 or 6, where 5 is awarded if the individual has had at least one year of normal weight, regular periods (in the case of AN), and no eating disorder behaviour.

5.2.2.2 Age of onset

Age of onset is the age of the participant in years, when they first met all signs and symptoms of the disorder.

5.2.2.3 Duration of illness

Duration of illness is taken as the time in years from the age of onset until the individual is considered recovered from the illness (i.e. severity becomes 5 or 6).

It should be noted that duration of illness for MDD is the time in years from onset of the first depressive episode, to the conclusion of the last depressive episode. Therefore if an individual had 4 separated MDE's of 6 months each over 10 years, their duration of illness would be recorded as 10 years. This is clarified as sometimes this measure for MDD details the duration of the longest MDE, rather than the time span between first and last MDE's.

5.2.3 SCID-II (OCPD only)

The obsessive-compulsive personality module of the Structured Clinical Interview for DSM-IV diagnosis of Axis II personality disorders was administered to ascertain level of OCPD. Nine questions targeting OCPD traits are asked, with each criterion scoring threshold if the participant's behaviour is pathological, pervasive and persistent. A subthreshold rating is given to criteria where two of the three "P's" are met. As with the SCID-I, both full (4 criteria scored threshold) and partial

diagnoses are given. Additionally, a continuous variable for level of OCPD was created by allocating a score of 1 to every threshold and 0.5 to every subthreshold question, creating a score between 0 and 9.

5.3   Self-report measures

5.3.1   Yale-Brown Obsessive-Compulsive Scale (Y-BOCS)

The Y-BOCS was originally developed as a clinican rated interview, consisting of a behaviour checklist and questions rating the severity of intrusions for first obsessive thoughts and then compulsive behaviours (Goodman et al., 1989a; Goodman et al., 1989b). Given that the Y-BOCS is a time intensive instrument to administer clinically, a self-report version has been developed and trialled (Steketee, Frost, & Bogart, 1996). This self-report version showed as good if not superior psychometric properties as the clinician-administered version. Checklist item examples are *"Concern or disgust with bodily waste and secretions"* and *"Checking locks, stove, appliances, faucets etc"*. Five questions regarding time occupied, interference, distress, effort to resist, and control are rated on a 5 point likert scale from 0 (none) to 4 (extreme) are completed after the checklist, regarding first principal obsessive thoughts and then principal compulsive behaviours. These questions are then totalled to give a score out of 20 for obsessions and compulsions, yielding a total Y-BOCS score out of 40 (obsessions + compulsions). The Y-BOCS is a sensitive instrument in the eating disorder population (Speranza et al., 2001), and has been used in a large-scale collaborative study of eating disorders (Kaye et al., 2008).

5.3.2   Yale-Brown-Cornell Eating Disorders Scale (YBC-EDS)

The YBC-EDS (Mazure, Halmi, Sunday, Romano, & Einhord, 1994) follows the same format as the Y-BOCS (outlined above) however items relating to thoughts and behaviours are specifically food, weight and shape related. Checklist item examples include *"Need to cut each piece of food into a specific size"*, and *"Fear of wearing tight or loose fitting clothing"*. Four follow-up questions regarding time occupied, interference, distress and control are rated on the same scale as that for the YBOCS, to give ED preoccupations and ED rituals scores out of 16, and a total YBC-EDS score out of 32. The YBC-EDS effectively distinguishes between current, recovered, and women with no ED (Sunday & Halmi, 2000), as well as full and partial syndrome AN and BN (Crow, Agras, Halmi, Mitchell, & Kraemer, 2002).

5.3.3   Hospital Anxiety and Depression scale (HADS)

The HADS (Zigmond & Snaith, 1983) is a 14 item questionnaire alternatively addressing anxiety and depressive symptomology over the past week. Each question has its own specific set of response anchors for the 4-point likert scale. For example, in response to the question *"I still enjoy the things I used to enjoy"*, participants could underline *Definitely as much* (to score 0), *Not quite so much* (to score 1), *Only a little* (to score 2) or *Hardly at all* (to score 3). Questions 2, 4, 7, 9, 12 & 14 are reverse scored. A higher score on each subscale is indicative of higher depressive mood or higher anxiety. The HADS has been utilised amongst diverse clinical groups, including

The HADS was one of the measures employed in this study to screen healthy control participants. Following recommendations from Zigmond and Snaith, participants with a depression or anxiety subscale score higher than 8 (possible anxiety/ depression) were excluded from the study.

5.3.4   Obsessive-compulsive Inventory-Revised (OCI-R)

The OCI-R (Foa et al., 2002) is an updated, shortened version of the Obsessive Compulsive Inventory (OCI; Foa, Kozac, Salkovskis, Coles & Amir, 1998), designed for use as a screening instrument to measure levels of obsessive or compulsive behaviours in subclinical individuals. Participants rate the 18 items on 5-point likert scale from 0 (*Not at all*) to 4 (*Extremely*). The scale consists of six subscales with three items each; Checking, Washing, Ordering, Obsessing, Neutralising, Hording . Item examples include *"I check things more often that usual"* and *"I need things to be arranged in a particular order"*. There are no counterbalanced items. The OCI-R has strong psychometric properties among both anxious (α=0.90) and non-anxious (α= 0.88) populations (Foa et al., 2002; Hajcak, Huppert, Simons, & Foa, 2004). The OCI-R was found to be free of socially desirable responding in a female undergraduate sample in New Zealand (Roberts & Wilson, 2008).

The OCI-R was one of the measures employed in this study to screen healthy control participants. Following recommendations from Foa and colleagues, any control participants scoring over 21 were screened out.

5.3.5   Rosenberg Self-esteem Scale (RSE)

The RSE (Rosenberg, 1965) is a short, 10-item counterbalanced measure of general self-esteem that has been widely used over the last four decades. Participants

rate each item on a 4-point likert scale from *Strongly Agree* to *Strongly Disagree*. Item examples include "*On the whole, I am satisfied with myself*" and "*I feel I do not have much to be proud of*". Items 2, 5, 6, 8 & 9 are reverse scored so that the higher the total scales score, the higher the individuals' self-esteem. Internal reliability is high, ranging from 0.77 – 0.88 across various samples (Rosenberg, 1986).

5.3.6   *Frost Multidimensional Perfectionism Scale (MPS)*

The MPS (Frost, Marten, Lahart, & Rosenblate, 1990) is a 35 item unbalanced scale, originally developed to measure an individuals level of perfectionism across six dimensions; Concern over mistakes (9 items), Personal standards (7 items), Parental expectation (5 items), Parental criticism (4 items), Doubts about actions (4 items), and Organisation (6 items). For the purposes of this study and based on previous large scale investigations (Price-foundation and GAN studies), the parental expectation and parental criticism subscales were removed. Item examples include *"If I fail partly, it is as bad as being a complete failure"*, and *"Only outstanding performance is acceptable in my family"*. Responses are measured on a 5-point likert scale from *Strongly Disagree* (score of 1) to *Strongly Agree* (score of 5). Research subsequent to the development of the scale has shown some controversy surrounding factorial stability of the scale. Reformulation using principal components analysis revealed that four subscales (Concerns over mistakes and doubts α=0.88; Parental expectations and criticism α=0.89, Personal standards α=0.78, and Organisation α=0.86) significantly increased the stability of the scale (Stober, 1998), suggesting an over extraction of factors at the initial scale development. In line with the large studies in the ED population employing this measure, the four original subscales (excluding parental subscales) will be used.

5.3.7   *Childhood Retrospective Perfectionism Questionnaire (ChiRP)*

The ChiRP (Southgate, Tchanturia, Collier, & Treasure, 2008a) is a self-report questionnaire developed from the EATATE interview part B (Brecelj-Anderluh et al., 2003). Using a yes/no format, the questionnaire addresses 20 obsessive-compulsive traits, asking whether each trait was present in the individuals life before the age of 12. Three domains are measured; childhood perfectionism (e.g. *"Did you always strive for the best mark at school or get upset if you were not always top of the class?"*), rigidity (e.g. *"Could you cope with changing plans at short notice?"*), and drive for order and symmetry (e.g. *"Did you like to make sure that everything was "just so" and in it's proper place?"*). One item *"Could you cope*

with changing plans at short notice" is reverse scored. The ChiRP is a new measure and so has not been extensively validated, however initial psychometric analysis indicate it has adequate test-retest, inter-rater reliability and concurrent validity (Southgate et al., 2008a). See Appendix 5 for a copy of the questionnaire.

5.3.8 Thinking Styles Questionnaire (TSQ)

The TSQ is a seven-item self-report questionnaire designed to measure an individuals style of thinking (Powell & Malia, 2003). Items are rated on a likert scale from *Strongly disagree* (1) to *Strongly agree* (6). Item examples include *"I have problems planning realistic goals, and working out the steps to achieve those goals"*, and *"I have difficulty planning and organising"*. See Appendix 5 for a copy of the questionnaire.

5.3.9 Cognitive Flexibility Questionnaire (CFS)

The CFS is a 12-item self-report questionnaire designed to measure an individuals awareness, willingness, and self-efficacy with regard to their own cognitive flexibility (Martin & Rubin, 1995). Participants rate their responses on a six-point likert scale from *Strongly disagree* (1) to *Strongly agree* (6). Item examples include *"I avoid new and unusual situations"* and *"My behaviour is a result of conscious decisions that I make"*. The CFS shows good construct validity, with scores from participant friends correlating well with individual self-report (Martin & Anderson, 1998). See Appendix 5 for a copy of the questionnaire.

5.3.10 Eating Attitudes Test-26 (EAT-26)

The EAT-26 (Garner, Olmsted, Bohr, & Garfinkel, 1982) is a shortened version of the 40-item Eating Attitudes Test (EAT; Garner & Garfinkel, 1979), designed to indicate disordered eating patterns among subclinical populations. It is not intended for use as a diagnostic tool. The 26-item scale is highly correlated with the original measure (Garner et al., 1982). Participants indicate their response to each item as *Always, Usually, Often, Sometimes, Rarely* or *Never.* Responses are then scored as follows; *Always* (3), *Usually* (2), *Often* (1), *Sometimes/Rarely/Never* (0). Item number 25 (*"I enjoy trying new rich foods"*) is reverse scored. Items comprise three subscales; dieting, bulimia, and oral control. Item examples include *"I avoid foods with sugar in them"* and *"I feel extremely guilty after eating"*. The EAT-26 has been widely used as a screening measure for eating disorders, and exhibits high internal reliability in both clinical ($\alpha = 0.90$) and subclinical ($\alpha = 0.83$) populations

(Garner et al., 1982). Moreover, Tilgner et al. (2004) found that the EAT-26 was free from social desirability bias in a sample of high school girls.

The EAT-26 was one of the measures used to screen HC participants. Following recommendations by Garner et al., participants with a score above 20 (after recoding) were excluded from the control group.

5.4   Ethical approval

This study had ethical approval from the Institute of Psychiatry Research Ethics Committee (ref 020/05, and 125/01). During the study, an updated ethics application designed to include both this project and further research resulting from it was submitted and approved (CREC/07/08-67). No ethical issues arose during the project.

5.5   Procedure

5.5.1   Recruitment

Participants were recruited through a large variety of sources both from the community and through the South London & Maudsley NHS trust Eating Disorder Service.

5.5.1.1   Clinical groups

Clinical participants were first recruited via a mailout to members of the Eating Disorders Unit volunteer database. Volunteers receive a bi-annual mailout, which includes a newsletter updating them on research progress in the unit, along with flyers of current studies that are actively seeking participants (see Appendix 6. for flyer). Participants were also recruited through advertisements on the research pages of our website www.eatingresearch.com and that of the eating disorder charity "b-eat" www.b-eat.co.uk. During the recruitment timeframe of this study, members of our unit were invited to participate in a number of media engagements such as interviews on BBC Radio 4, BBC One's The Breakfast Show, and for newspapers such as The Evening Standard, The Post, and The Times. A large number of responses to these events were received and any suitable candidates were referred to the study. Finally any suitable patients coming through either the inpatient or outpatient wards of the South London & Maudsley NHS trust were invited to participate in the research.

5.5.1.2   Healthy control group

HC were recruited from a wide range of domains across Greater London. In the first instance, a circular email was sent to King's College London staff and students. Advertisements for participants were posted in libraries, local churches, café's, shop windows and public facilities. Permission was obtained from Transport for London to hand out flyers at London Bridge station for a day. The Institute of Psychiatry's healthy volunteer database "MindSearch" was also used to access willing participants. Electronic advertisements were posted on popular Internet sites such as gumtree and facebook. Study controls were recruited from all walks of life, in order to limit cluster sampling and over sampling of easy to reach groups (e.g. university students).

5.5.2   Inclusion/exclusion criteria

5.5.2.1   Clinical groups

For inclusion in the clinical groups, participants must have met lifetime criteria for a DSM-IV diagnosis of AN or BN, and have a current age between 16-60. Unaffected sisters must have a sister with lifetime AN or BN, and a current age between 16-60. Participants meeting criteria for EDNOS only were excluded. Participants were not excluded based on any other factors such as co-morbid diagnosis (e.g. depression or dyslexia) or current medication.

5.5.2.2   Healthy control group

HC participants were carefully screened in a two-stage process. Firstly a screen was conducted via telephone or email, to check for personal or family history of (diagnosed) mental illness, traumatic brain injury, healthy BMI (17.5 - 25) and ethnicity (white European decent) to match that of the clinical group. If participants fell outside any of these criteria they were excluded from the study. Secondly, completed self-report questionnaires were individually assessed for measures investigating disordered eating behaviour (EAT-26), obsessive-compulsive behaviour (OCI-R), and anxiety and depression (HADS). Participants scoring above the cut-off on any of these measures were excluded (see 5.3 for exact cut-off scores). Those passing both screening stages were invited to participate in the study.

5.5.3   Session protocol

Participants that had volunteered for the project were sent an information and consent pack in the post, along with the self-report measures. An addressed, freepost envelope was included in the pack for them to return the consent form and

questionnaires. The information sheet clearly stated that participation in the project was voluntary, and that participants were free to pull out at any time without having to give a reason (see Appendix 7 for a copy of the information sheet)

The researcher liaised with participants and booked an appointment time. In the vast majority of cases this appointment was held at the Eating Disorders Research Unit at Guy's hospital, in a small quiet assessment room. In some cases the researcher travelled to participants' homes or workplaces, or if the participant was a current inpatient the researcher travelled to the inpatient ward. In these cases every attempt was made to replicate the conditions of the testing environment at Guy's (laptop position, undisturbed setting etc).

5.5.3.1 Clinical groups

Sessions began with the SCID extended module H. All participants including those participating as unaffected sisters and those who already had a clinical diagnosis were assessed for current and lifetime ED. Participants were then assessed with the full SCID-I, and SCID-II OCPD section. The researcher also briefly checked self-report responses to the Y-BOCS and YBC-EDS, checking the item ratings the participant had given in response to the checklists. If necessary the researcher revisited these measures with the participant, as it was found that participants were often confused with the time frame or by the follow-up questions themselves. The interviews took anywhere from 15 minutes to 1.5 hours to complete with each participant.

Participants were then administered the Neuropsychology battery in the following order: TMT, WCST, Brixton, ROCF copy, GEFT, CatBat, Haptic, RCF recall. If a participant took a particularly long time on the GEFT (likely due to a large number of time-out errors), the RCF recall was administered second to last, and the Haptic task moved to last place. This was to maintain consistency in the 20-minute delay between the ROCF copy and recall trials. This change was made at the discretion of the researcher.

Following the neuropsychological assessment the researcher debriefed the participant, explaining each of the tasks and how they related to the two aspects of "thinking style" being assessed. This was done in an informal way, and it was found helpful to draw a diagram where the researcher would plot the approximate performance on each of the aspects of thinking style (i.e. more toward the detail or global end, and more toward flexible or rigid end).

Finally, the researcher drew two 8ml tubes of blood for DNA analysis. In some cases, for example when the participant had an appointment late in the day, the blood was drawn at the beginning of the session to ensure the sample could be sent to the IoP genetics laboratory the same day. Blood samples were packaged appropriately and either posted first class or hand delivered to the genetics lab.

Participants were thanked for their time and offered £30 as compensation for their participation, along with any reasonable travel costs incurred (usually no more than an additional £20). Each session took between 1.5-2 hours to complete.

5.5.3.2 Healthy control group

Session protocol for the HC group consisted of the same neuropsychological assessment outlined above and blood sample for DNA. No interviews were undertaken with control participants. As some of the self-report measures were used in the second stage of screening, these had been completed and posted back to the researcher for screening before the session was booked. Participants were offered £20 to thank them for their time. The session took approximately an hour.

5.6   Statistical Methods

5.6.1   Power Analysis

Power calculations were derived using nQuery software (see Table 7). Calculations for most set-shifting and coherence tasks employed previously published data from our group (Tchanturia et al., 2004a; Lopez et al., 2008b). As this is the first time our group has used the WCST and GEFT, power calculations for these tasks were calculated using published data from external groups (Steinglass et al., 2006). It should be noted that different administration methodology was adopted for the GEFT paper, where an upper time limit of 180 seconds was applied (compared to the current studies 60 seconds), and mean time rather than median time was calculated. Based on these previous findings, nQuery predicted that a sample size of 40 in each group (sample size 19-40 depending on the task) will have 80% power to detect a difference in means, using a two group t-test with a 0.05 two-sided significance level.

5.6.2   Management of outliers

The main outcome variable from each task was chosen to assess for outliers as follows; TMT B-A time, WCST perseverative errors, Brixton errors, CatBat B-C time, Haptic perseverations, ROCF central coherence index, and GEFT median time.

Table 7: Power analysis by neuropsychological task

Task	Source	Power	N required
TMT	Tchanturia et al. (2004a)	80%	21
WCST perseverative errors	Steinglass et al. (2006)	80%	22
Brixton Task	Tchanturia et al. (2004a)	80%	26
CatBat Task	Tchanturia et al. (2004a)	80%	40
Haptic Illusion	Tchanturia et al. (2004a)	80%	30
ROCF central coherence index	Lopez et al. (2008a)	80%	19
GEFT	Tokley & Kemps (2007)	80%	29

TMT Trail Making Test; WCST Wisconsin Card Sorting Test; ROCF Rey-Osterrieth Complex Figure; GEFT Group Embedded Figure Test.

Outliers were determined by inspecting boxplots of each outcome variable split by diagnosis (current AN; current BN; recovered AN; AN healthy sister: BN healthy sister; HC). See Appendix 8 for boxplots from which outliers were determined. It was decided that outlying data points would be removed from the dataset. During administration, it was observed that a number of participants simply gave up on a task. In such cases, the participant's results reflected a lack of motivation rather than for example poor flexibility on the WCST. In an attempt to eliminate these unrepresentative results from the dataset, outliers were removed. The limitations of this approach are acknowledged. In particular, it is likely that this also resulted in the deletion of cases where, following the above example, an individual was actively attempting to complete the WCST however failed to grasp the sorting rule. In order to reduce this possibility, outliers were identified within the context of each clinical groups performance.

Given the difficulty in obtaining large sample sizes of neuropsychological data, excluding outliers on a case-by-case basis (i.e. deleting all the data from a participant where an outlier is present on one task) was not feasible due to the large number of cases that would be lost. Therefore outlying data points were deleted on a task-by-task basis. Across tasks, between 0% and 8.5% of data points were excluded (see Table 8). Using this method, 4.2% of all possible data points were deemed outliers and excluded from the analysis (79 out of a possible 1890 data points).

5.6.3   *Procedure for normality assessment*

Each outcome variable of interest for each neuropsychological task is assessed to determine whether data shows a normal (Gaussian) or non-normal distribution. This is assessed individually across each separate analysis. Normality is determined by weighing evidence across three methods: Firstly, boxplots split by diagnosis (HC, AN, BN, recovered AN, AN sister, and/or BN sister) are visually assessed to investigate how comparable the data spread is across groups. Skewedness and kurtosis are taken into account. Where groups depict a dissimilar spread, this is taken as an indication of non-normal distribution. Shapiro-Wilks test for normality is also be employed, where a significant test is taken as an indication of non-normal distribution. Thirdly, the standard deviation of each outcome is assessed. If the standard deviation of one group is more than twice that of another, this is taken as an indication of non-normal distribution. Error count outcome variables (TMT errors, CatBat errors, GEFT time-out fails) are conceptually hypothesised to be

Table 8: Frequency of outlier deletion by group and task

	HC	AN	BN	rcAN	ANsis	BNsis	Total outliers (by task)	
TMT B-A	2	0	0	6	2	1	11	4.1%
WCST perseverative errors	7	3	4	2	4	3	23	8.5%
Brixton Test	0	0	0	0	0	0	0	0%
CatBat Task B-C	6	6	2	2	5	1	22	8.1%
Haptic Task	0	0	0	0	0	0	0	0%
ROCF central coherence index	7	0	0	3	0	0	10	3.7%
GEFT median time	3	3	3	3	1	1	14	5.2%
Total outliers (by group)	25	12	9	16	12	6		
	4.1%	2.5%	4.3%	7.6%	5.7%	4.3%		

HC Healthy Control; AN Anorexia Nervosa; BN Bulimia Nervosa; rcAN Recovered Anorexia Nervosa; ANsis Anorexia Nervosa Unaffected sister; BNsis Bulimia Nervosa Unaffected sister; TMT Trail Making Test; WCST Wisconsin Card Sorting Test; ROCF Rey-Osterrieth Complex Figure; GEFT Group Embedded Figure Test.
Note: Calculations do not take into account missing data

non-normal, as are variables where a time cut-off or administration maximum is imposed (GEFT median time, WCST categories completed, Haptic perseverations). Outcome of normality assessment is presented in each relevant methodology section (e.g. 6.2.1.2).

5.6.4 Selection of covariates

Psychological factors such as depression and anxiety were not considered as covariates given that their presence is inherent in an ED. Therefore, only demographic factors were considered. Participants were matched by selection on gender and ethnicity. Age is a potential covariate given it can differ significantly between clinical and control groups (see following chapters). Where age differs significantly between groups, correlates with outcome and where parametric statistics are used, age will be employed as a covariate. This will be determined analysis by analysis. Like age, years of education can differ significantly between groups (see following chapters). However the presence of an ED often means that education is interrupted, and lower years of education (by approximately 2 years) yet comparable estimates of IQ have been previously reported when comparing ED with HC participants (Lopez et al., 2008b). Therefore years of education will not be controlled for.

5.6.5 Statistical analysis

A wide range of statistical analyses will be employed in this thesis, ranging from independent-samples t-tests and analysis of covariance (ANCOVA) to Wilcoxon Signed Ranks Tests. The decision as to which test will be employed will be made independently for each analysis, depending on planned group comparisons and normality of the data. Therefore specific details as to which tests are employed will be detailed in the method section of each empirical chapter. Hochberg's correction will be applied to neuropsychological and self-report analyses to adjust for multiple testing. Where a result is no longer significant following the correction, this will be noted in the data tables.

Cohen's d effect sizes are calculated for all self-report and neuropsychological comparisons. Differences are defined as negligible ($\geq$ -0.15 and < 0.15), small ($\geq$ 0.15 and < 0.40), medium ($\geq$ 0.40 and < 0.75), large ($\geq$ 0.75 and < 1.10), very large, ($\geq$ 1.10 and < 1.45) and huge ($\geq$ 1.45).

All analyses will be carried out using the Statistical Package for Social Sciences (SPSS) for Apple Mac version 16.0. Rigorous checks were made on the

data both pre and post entry into the SPSS spreadsheet. As appropriate, decisions on data analysis were made in consultation with Dr Sophie Barthel and Dr Daniel Stahl, statisticians within the Biostatics department of the Institute of Psychiatry.

6   Study 1- The endophenotype is associated with illness in the population

6.1   Background

Aspects of neuropsychological profile are of particular relevance to the search for endophenotypes (see chapter 3.2). In order for a trait to be considered a candidate endophenotype, it must first be observed at a higher rate within the population under investigation compared to the general population. Neurocognitive deficits in executive functioning, visuo-spatial ability, attention, and memory amongst others have been observed in the eating disorder (ED) population (Lena, Fiocco, & Leyenaar, 2004). Such difficulties cannot be explained by starvation alone given their presence in both anorexia (AN) and bulimia nervosa (BN), or by levels of general intelligence as those with ED tend to display comparable levels of IQ compared to matched control women (Gillberg et al., 1996).

This thesis is focussed on the assessment of two specific neurocognitive traits as endophenotypes of ED: poor cognitive set-shifting (an inflexible or rigid processing style) and weak coherence (a bias toward local processing, often to the exclusion of context). A recent meta-analysis (see chapter 4) highlighted empirical support for poor cognitive flexibility in ED across 15 studies from a number of different research groups employing various neuropsychological measures (Roberts, Tchanturia, Stahl, Southgate, & Treasure, 2007b). Less evidence for superior local processing/poor global integration exists simply because the concept of weak coherence has only recently been applied to ED (Lopez et al., 2008b; 2008d). A recent systematic review of tasks tapping weak coherence has highlighted consistent findings where methodology has allowed for detail focus to be measured in the absence of confounding variables such as working memory (Lopez et al., 2008c).

The present chapter (Study 1) will for the first time examine both set-shifting and weak coherence in the same individuals, and add to the set-shifting battery the Wisconsin Card Sorting Test (a compurerised task used as standard in the investigation of this trait other psychiatric conditions). Study 1 aims to replicate previous findings for both neurocognitive biases in a new cohort of women with and without an ED, in order to address the first criterion of an endophenotype as outlined by Gottesman and Gould (2003) for each trait: "the endophenotype is associated with illness in the population". In past research, sample size limitations have meant that diagnostic subtypes have been collapsed when assessing neuropsychological

performance, for example combining restricting (ANR) and binge/purging (ANBP) subtypes of AN, or combining AN and BN groups. The scope of the current study has allowed for subtyping within the current ED group based on both traditional diagnostic criteria and on illness symptomology, adding depth to the replication.

6.1.1 Hypotheses

It is hypothesised that ED groups will show 1) poor cognitive flexibility compared to the healthy control (HC) group, and 2) a bias toward detailed processing compared to the HC group. Secondly, it is hypothesised that these biases will be stronger in those with AN compared to BN.

6.2 Method and Results

The general methodology was outlined in Chapter 5. This combined method and results section will present findings split into four sections. The first three sections will assess neuropsychological performance across different clinical classifications; 1) a transdiagnostic assessment of current ED compared to HC, 2) traditional DSM-IV classifications of current AN (ANR & ANBP) and current bulimia nervosa (BN) compared to HC, and 3) assessment based on lifetime phenotype, with pure restrictors (ANR), those displaying a mix of AN and BN behaviours (ANBN), and 'pure' BN compared to HC. The final section will create a composite variable based on task performance to assess the relationship between neuropsychology and illness-associated variables. Investigating results from these different viewpoints will aid further understanding of set-shifting and weak coherence as they relate to the different presentations of ED.

6.2.1 Method Analysis 1: Transdiagnostic split

This first stage of the analysis adopts a transdiagnostic approach, comparing women with current AN or BN with women with no personal or family history of ED. This analysis will test the broad hypothesis of poor shifting and weak coherence across the ED spectrum. The measures administered and procedure followed were outlined in chapter 5.

6.2.1.1 Participants

Participants were 98 women with ED where current severity rating ranged between severe and partial remission (25 or 25.5% severe; 28 or 28.6% moderate; 24 or 24.5% mild; 21 or 21.4% partial remission), and 88 HC women.

6.2.1.2   Statistical methods

Neuropsychological variables were assessed for normality using three criteria (as detailed in 5.6.3); Boxplot assessment, Shapiro-Wilk test, and assessment of standard deviations across groups. Seven of the 17 variables were deemed normally distributed (TMT shift time; TMT B-A; Brixton task; CatBat shift time; CatBat B-C; ROCF copy accuracy; ROCF recall accuracy). Differences will therefore analysed using independent-samples t-tests, with means and standard deviations (in parentheses) presented in the tables. Age differed significantly between groups and correlated with both TMT shift time ($r(182)=0.16$, $p=0.04$) and ROCF recall accuracy ($r(178)=-0.24$, $p=0.001$). Therefore these two variables will be assessed using ANCOVA with age as a covariate. Mann-Whitney U tests will be used to explore group differences for all other variables. For these variables, descriptive statistics presented are the median, with 25th and 75th quartiles (in parentheses).

Self-report measures were largely normally distributed (HADS anxiety; Rosenberg self-esteem; Frost perfectionism; Thinking styles scale; Cognitive flexibility scale) and so will be analysed with independent-samples t-tests. HADS depression, OCI-R and CHiRP were non-normally distributed and therefore will be analysed with Mann-Whitney U tests.

6.2.2   Results Analysis 1: Transdiagnostic split

6.2.2.1   Demographic and Clinical Features

All participants were female and of white European decent. The HC group was on average older and had more years of education than the clinical group (see Table 9). As expected given the high proportion of women with a diagnosis of AN in the current ED group (69.4%), a significantly lower BMI was found in clinical group. Women with ED were on average in a moderate to mild stage of the illness with an eight-year duration of illness, however substantial variance within the group was observed with regard to illness duration. Age of onset was 17 years. The scores on the YBC-EDS rated for the worst period of the illness showed a moderate-severe symptom level with minimal variation in scores, indicating that all participants had a similar illness severity at its worst.

Self-report Clinical Features: Table 10 details self-report clinical features of both groups. Consistent significant differences emerged between groups on self-report questionnaires, where the ED group showed high scores on all questionnaires. For some measures this was largely influenced by HC group screening.

Table 9: Study 1.1 Demographic and clinical features

	ED	HC	Test statistics	
	(n=98)	(n=88)	t	p-value
Age	25.17 (6.99)	28.43 (8.47)	2.87	<0.01**
Years of Education [eq]	15.66 (2.65)	16.76 (1.98)	3.09	<0.01**
BMI (current) [eq]	19.07 (3.20)	22.07 (1.79)	7.97	<0.001**
BMI (lowest)	15.50 (4.13)	-		
BMI (highest)	22.72 (3.42)	-		
Current Severity	2.42 (1.09)	-		
Age of ED Onset	16.78 (4.03)	-		
Duration of Illness	8.14 (5.99)	-		

ED Eating Disorder; HC Healthy Control; BMI Body mass index
[eq] Equal variances not assumed (Levene's test for equality of variance <0.05)
** Comparison significant at 0.01 level

Table 10: Study 1.1 Self-report clinical features

	ED	HC	Test statistic			
	(n=94)	(n=86)	t	MW	p-value	Cohen's d [2]
HADS anxiety [eq]	12.09 (4.72)	4.20 (2.32)	-14.45	-	<0.001	2.09**
HADS depression [1]	6 (3.38-10)	1 (0-2)	-	605.00	<0.001	2.21**
OCI-R total [1]	18 (10-29)	5.5 (2.25-10)	-	1375.00	<0.001	1.41**
Rosenberg self-esteem [eq]	11.63 (5.71)	23.52 (3.96)	16.43	-	<0.001	-2.40**
Frost Perfectionism	96.57 (15.99)	72.95 (15.22) [3]	-6.18	-	<0.001	1.49**
CHiRP total [1]	7 (4-11)	2.5 (1-4) [3]	-	432.00	<0.001	1.05*
Y-BOCS	3.5 (0-20.25)	-				
YBC-EDS	23.92 (5.57)	-				
Thinking Styles [eq]	25.38 (7.20)	14.66 (4.89)	-11.86	-	<0.001	-1.73**
Cognitive Flexibility [eq]	47.28 (9.54)	60.34 (5.96)	11.05	-	<0.001	1.63**

ED Eating Disorder; HC Healthy Control; MW Mann-Whitney U Test; HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire; YBC-EDS Yale-Brown-Cornell Eating Disorder Scale; Y-BOCS Yale-Brown Obsessive-Compulsive Scale.
[eq] Equal variances not assumed (Levene's test for equality of variance <0.05)
[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses
[2] Cohen's d effect size comparisons
[3] HC data collected from a subset of participants (n=22)
* Comparison significant at 0.05 level
** Comparison significant at 0.01 level

For example, scores for HADS anxiety and depression and OCI-R were low as HC participants were excluded from the study if their scores were high (see 5.5.2.2). HC were also screened for ED behaviours, therefore scores on the EAT-26 were low (EAT total 5.07, SD=4.15; dieting subscale M=3.31, SD=3.42; bulimia subscale M=0.35, SD=0.72; oral control subscale M=1.41, SD=1.54).

Self-report Cognitive Style: Very large effect sizes were seen on both the cognitive flexibility scale (CFS) and the thinking styles questionnaire (TSQ). As the questionnaires load in opposite directions, both results indicate that women with current ED perceived themselves as having poorer cognitive flexibility than HC.

6.2.2.2   Comorbidity

Nearly 75% of the sample met full or partial criteria for depression, where on average current severity was partial remission (see Table 11). Obsessive-compulsive disorder (OCD) was next most highly endorsed, with almost half of participants receiving a diagnosis. OCD age of onset was on average 14 years, indicating that OCD preceded ED onset by nearly three years. The most frequently endorsed subtypes were ordering (43.1%) and washing (29.2%). Full and partial criteria obsessive-compulsive personality disorder (OCPD) was present in one third of participants. Social phobia, specific phobia and generalised anxiety disorder (GAD) diagnoses were on average moderate to mild in current severity, also showing similar ages of onset (approx. 12 years) and illness duration (approx. 15 years). Body dysmorphic disorder (BDD) was only endorsed by one participant at sub-threshold level, only with one year of illness duration. Self-harming behaviours were present in 41 or 43.6% of those with ED.

Approximately 25% of clinical participants met criteria for alcohol or substance disorders. Those with substance abuse/dependence were largely recovered at the time of assessment, where those with alcohol abuse/dependence were in partial remission. Age of onset was similar across diagnoses, however more variance in duration of illness was observed in the alcohol group.

6.2.2.3   Set-shifting results

Descriptive statistics, data analysis and effect size results for set-shifting variables (ED vs HC) are presented in Table 12.

Trail Making Test (TMT): A significant difference was found between groups on the TMT raw shift time, where those with a diagnosis of ED took longer on the shift trial than HC. Age was a significant covariate (p<0.01). Results for the balanced

Table 11: Study 1.1 Comorbid psychiatric diagnoses for ED

	Full	Partial	Diagnosis[1]	Diagnostic details			
				N	Severity[2]	AOO (yrs)	DOI (yrs)
Anxiety Disorders							
OCD	35 (37.6%)	10 (10.8%)	45 (48.4%)	37	3.33 (1.38)	13.78 (5.48)	8.09 (6.79)
OCPD	15 (16.1%)	17 (18.3%)	32 (34.4%)	-			
Panic Disorder	19 (19.6%)	11 (11.3%)	30 (30.9%)	26	4.40 (1.21)	21.04 (6.88)	4.72 (5.14)
Social Phobia	32 (33.0%)	7 (7.2%)	39 (40.2%)	39	2.64 (1.18)	10.26 (6.31)	16.21 (11.77)
Specific Phobia	13 (13.4%)	9 (9.3%)	22 (22.7%)	22	2.41 (1.01)	11.09 (7.33)	14.18 (7.54)
PTSD	8 (8.3%)	3 (3.1%)	12 (12.4%)	11	3.36 (1.36)	18.55 (9.14)	6.40 (7.25)
GAD	14 (14.4%)	3 (3.1%)	17 (17.5%)	17	2.29 (0.47)	12.12 (7.40)	14.29 (8.87)
BDD	0	1 (1.0%)	1 (1.0%)	1	3.00 (-)	43.00 (-)	1.00 (-)
Mood Disorders							
MDD	60 (61.9%)	11 (11.3%)	71 (73.2%)	63	4.06 (1.41)	16.62 (5.00)	5.58 (6.45) [3]
Bipolar	3 (3.1%)	2 (2.1%)	5 (5.2%)	5	3.80 (0.45)	16.80 (1.48)	4.70 (3.93)
Dysthymia	2 (2.1%)	1 (1.0%)	3 (3.1%)	2	5.50 (0.71)	15.50 (0.71)	2.50 (0.71)
Substance Disorders							
Alcohol A/D	24 (24.7%)	9 (4.1%)	33 (28.9%)	26	3.96 (1.70)	20.37 (5.71)	3.71 (5.96)
Sub. A/D	20 (20.6%)	1 (1.0%)	21 (21.6%)	14	4.71 (1.53)	19.65 (5.12)	5.00 (3.82)

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder; Alcohol A/D Alcohol abuse/dependence; Sub. A/D Substance abuse/dependence.

[1]Diagnosis indicates pooled data from full and partial (1 criterion short) diagnosis. Severity, AOO and DOI details are for pooled data.

[2]Severity rated on 6-point scale from 1 (severe) to 6 (prior history), where 5 and 6 indicate fully recovered (>1 year)

[3]Period of time over which depressive episode/s occurred.

Table 12: Study 1.1 Set-shifting descriptive statistics

	ED	HC	Test Statistic			
	(n=90)	(n=78)	F/t	MW	p-value	Cohen's d [2]
TMT shift time (B) [3]	30.63 (8.13)	28.08 (6.92)	6.13	-	<0.01	0.34**
TMT B-A	10.56 (6.81)	8.89 (6.31)	-1.69	-	0.09	0.25
TMT errors (shift) [1]	0 (0-1)	0 (0-1)	-	3490.0	0.18	0.20
WCST Perseverative errors [1]	9 (6-17.5)	7 (5.75-9)	-	2462.0	0.001	0.54**
WCST Categories completed [1]	6 (6-6)	6 (6-6)	-	2728.5	<0.001	0.71**
Brixton errors	11.31 (4.10)	10.01 (4.21)	-2.12	-	0.04 [£]	0.31*
CatBat shift time (Bat)	31.16 (9.82)	29.08 (11.02)	-1.31	-	0.19	0.20
CatBat B-C	9.60 (7.47)	8.38 (7.58)	-1.06	-	0.29	0.16
CatBat errors (shift) [1]	0 (0-1)	0 (0-1)	-	3504.0	0.52	0.10
Haptic perseverations [1]	16 (10-30)	13 (7-21.75)	-	3440.0	0.02 [£]	0.34*

ED Eating Disorder; HC Healthy Control; MW Mann-Whitney U Test; TMT Trail Making Test; WCST Wisconsin Card Sorting Test

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons

[3] Age run as covariate, therefore F statistic presented

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

[£] Comparison no longer significant after Hochberg correction

B-A variable (shift trial time less control trial time) trended in the same direction however fell short of significance. No notable difference on number of errors was seen across the groups.

Wisconsin Card Sorting Test (WCST): A significant difference was found between ED and HC groups for WCST perseverative errors, where those with ED made significantly more errors (moderate effect size) with a notably wider range than the HC group. Although not obvious from the descriptive statistics, categories completed followed the same trend with the ED group completing significantly less categories than HC, with a large effect size. Inspection of the raw data revealed that despite both groups scoring at ceiling on the lower quartile, 22 ED participants completed between 0-5 categories compared to just one HC.

Brixton Task: A significant difference was found on the Brixton task, where the ED group made significantly more errors than HC. A small effect size was observed.

CatBat: No significant differences between groups were found for the CatBat task across either the raw shift time, balanced B-C (Bat time less Cat time) variable, or the number of errors. No notable trends are seen, with negligible to small effect sizes across outcomes.

Haptic: A significant difference was found on the Haptic task, where those with ED made more illusions than the HC group, with a small effect size. A large range is noted in the ED group, where the upper quartile is at ceiling.

6.2.2.4   Coherence results

Descriptive statistics, data analysis and effect size results for coherence specific outcome variables are presented in the first section of Table 13, with related outcome variables in the second section.

Group Embedded Figures Test (GEFT): A significant difference was found for both GEFT median time and time out errors, where those with ED were both faster and made less time out errors than the HC group, indicating a more detail focussed processing style.

Rey-Osterrieth Complex Figure (ROCF): A significant difference was found between groups for the central coherence index, where those with ED had a lower score indicating a more detail focussed processing style. A moderate to large effect size was seen. The ED group had a lower accuracy score than HC (small effect size) indicating that they remembered less of the figure after 20 minutes. This difference

Table 13: Study 1.1 Coherence descriptive statistics

| | ED | HC | Test Statistic | | | |
	(n=91)	(n=81)	t/F	MW	p-value	Cohen's d [2]
GEFT median [1]	7.72 (5.1-10.33)	8.85 (5.86-15.43)	-	3049.0	0.02	-0.35*
GEFT time out errors [1]	1 (0-2)	1 (1-2.75)	-	3073.5	0.02	0.36*
ROCF coherence index [1]	1.41 (0.94-1.61)	1.56 (1.41-1.69)	-	2387.5	<0.001	-0.73**
ROCF order [1]	2.17 (1.63-2.50)	2.45 (1.41-1.69)	-	2431.0	<0.001	-0.71**
ROCF style [1]	1.50 (1-1.67)	1.67 (1.50-1.83)	-	2512.0	<0.001	-0.67**
ROCF copy accuracy	28.64 (3.51)	29.31 (3.92)	1.44	-	0.23	-0.18
ROCF recall accuracy [3]	15.07 (5.53)	16.58 (4.87)	9.48	-	<0.001	-0.29**

ED Eating Disorder; HC Healthy Control; MW Mann-Whitney U Test; GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure
[1]Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses
[2] Cohen's d effect size comparisons
[3]Age run as covariate, therefore F statistic presented
* Comparison significant at 0.05 level
** Comparison significant at 0.01 level

did not reach significance. Age was a significant covariate (p<0.001). A significant, moderate, positive correlation was found between copy central coherence and recall accuracy across the whole sample, r(178)=0.42, p<0.001.

6.2.3 Method Analysis 2: Traditional diagnostic split

In order to more fully understand the findings presented in analysis 1, results from the same participants will be re-analysed split by the traditional AN/BN diagnostic categories of the DSM-IV. In general, investigations of neuropsychological profile in the ED population have used this classification system, as indeed the DSM-IV is the standard diagnostic tool used in psychiatry. The current study will investigate whether the cognitive profile of poor set-shifting and weak coherence found when comparing all ED to HC (method 1) persists amongst both AN and BN diagnoses when split accordingly.

6.2.3.1 Participants

Participants were those reported in analysis 1. For this analysis, the 98 women with ED were split into AN (n=68; 35 or 51.5% ANR, 33 or 48.5% ANB/P) and BN (n=30; 13 or 43.3% with a lifetime history of AN).

6.2.3.2 Statistical methods

Normality assessment of neuropsychological variables across the three groups was the same as that reported in the transdiagnostic split (method 1; see 6.2.1.2). The seven normally distributed variables will be analysed using ANOVA (ANCOVA with age as a covariate for TMT shift time and ROCF recall accuracy), with post-hoc Tukey or Least Significant Difference (LSD, for ANCOVA) tests for individual group comparisons. Kruskal-Wallis tests will be used to analyse overall group differences for non-normal data, with Mann-Whitney U tests employed to analyse post-hoc group comparisons. Normality of self-report measures was also the same as that reported in the transdiagnostic split.

6.2.4 Results Analysis 2: Traditional diagnostic split

6.2.4.1 Demographic and Clinical Features

As in the transdiagnostic split, the HC group trended toward being both older and having more years of education than both clinical groups, with this difference reaching significance compared to the AN group only (for descriptive statistics see Table 14). As expected, the AN group had a significantly lower body mass index

Table 14: Study 1.2 Demographic and Clinical Features

	AN (n=68)	BN (n-30)	HC (n=88)	Test statistic F	t	p-value
Age [a]	24.62 (7.03)	26.43 (6.84)	28.43 (8.47)	4.70	-	0.01**
Years of Education [a]	15.53 (2.56)	16.00 (2.89)	16.76 (1.98)	4.82	-	<0.01**
BMI (current) [a, b]	17.93 (2.60)	21.66 (2.94)	22.07 (1.79)	65.44	-	<0.001**
BMI (lowest)	14.55 (4.26)	17.63 (2.89)	-	-	3.60	0.001**
BMI (highest)	21.71 (3.06)	24.95 (3.15)	-	-	4.76	<0.001**
Current Severity	2.41 (1.07)	2.43 (1.17)	-	-	0.09	0.93
Age of ED Onset	16.74 (4.11)	16.87 (3.91)	-	-	0.15	0.88
Duration of Illness	7.67 (5.82)	9.22 (6.33)	-	-	1.81	0.24

AN Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control; BMI Body mass index
[eq] Equal variances not assumed (Levene's test for equality of variance <0.05)
[a] Significant difference (Tukey's HSD) between AN and HC
[b] Significant difference (Tukey's HSD) between AN and BN
** Comparison significant at 0.01 level

(BMI) compared to both BN and HC groups. No difference was found between BN and HC in terms of BMI. This healthy weight of the BN group indicates that, even though nearly half of the current BN group had a history of AN, their current state of illness was representative of normal weight BN rather than ANBP.

While the average lowest weight of the BN group was below an 18 BMI, this is accounted for by the AN history in the group. The reverse pattern is observed regarding highest weight ever, where the BN group had a highest BMI significantly higher than those with AN. No other differences in illness details (current severity, age of onset, duration of illness) were observed across clinical groups, however the AN group trended toward a shorter duration of illness (83% that of BN). Similarity of clinical features is particularly notable for results on the YBC where both the total score and preoccupations/rituals subscale scores did not differ, indicating similar illness severity at its worst time point across both AN and BN groups.

Self-report Clinical Features: Self-report details are presented in Table 15. As expected, scores in both ED groups differed significantly to the HC group across all measures, with very large effect sizes seen. Some differences were seen between clinical groups, namely the BN group had significantly higher depression and anxiety scores than the AN group.

Self-report cognitive style: A similar pattern was observed across both AN and BN groups in comparison to HC, where scores on the CFS and TSQ differed substantially between clinical and control groups. The BN group scored significantly higher on the TSQ than the AN group ($p<0.01$), with a moderate effect size.

6.2.4.2   Comorbidity

Comorbid diagnostic details for AN and BN groups can be found in Table 16 and Table 17. The most highly endorsed comorbid condition was Major Depressive Disorder, where approximately 75% of all clinical groups met threshold or subthreshold criteria. Those with AN showed a significantly shorter duration of depressed mood compared to BN ($t(62)=-2.31$, $p=0.02$). Little endorsement of Bipolar or Dysthymia was observed across the clinical groups.

OCD was next highest endorsed in the AN group, with approximately half of the sample meeting threshold or subthreshold criteria. Age of onset was low across both clinical groups at just under 14 years of age, indicating that OCD tendencies preceded the onset of the ED by approximately three years. The prominent subtype across AN and BN was ordering (see Table 18). While the AN group largely

Table 15: Study 1.2 Self-report clinical features

	AN	BN	HC	Test statistic			Cohen's d [2]	
	(n=66)	(n=29)	(n=88)	F/t	KW	p-value	AN	BN
HADS anxiety	11.32 (4.82)	13.83 (4.05)	4.20 (2.32)	108.83	-	<0.001	1.97**	3.39**
HADS depression [1]	6 (3.25-8)	7 (4-12.5)	1 (0-2)	-	100.98	<0.001	2.02**	2.01**
OCI-R total [1]	16 (7-27)	22 (14-31)	5.5 (2.25-10)	-	64.48	<0.001	1.15**	1.66**
Rosenberg self-esteem	11.82 (5.77)	11.21 (5.63)	23.51 (3.96)	130.54	-	<0.001	-2.43**	-2.78**
Frost Perfectionism	96.54 (17.70)	96.64 (14.52)	72.95 (15.22) [3]	18.91	-	<0.001	1.45**	1.57**
CHiRP total [1]	3 (1-4)	1 (0-3)	1 (0-2) [3]	-	22.52	<0.001	1.30**	0.90**
Y-BOCS	13.77 (12.95)	13.38 (12.34)	-	-0.14	-	0.89	-	
YBC-EDS	24.19 (4.93)	23.34 (6.81)	-	-0.68	-	0.50	-	
Thinking Styles	23.93 (6.28)	28.68 (8.15)	14.66 (4.89)	78.99	-	<0.001	1.67**	2.39**
Cognitive Flexibility	47.95 (9.16)	45.75 (10.37)	60.34 (5.96)	60.47	-	<0.001	-1.64**	-2.00**

AN Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control; HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire; YBC-EDS Yale-Brown-Cornell Eating Disorder Scale; Y-BOCS Yale-Brown Obsessive-Compulsive Scale.

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons with HC data

[3] HC data collected from a subset of participants (n=22)

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

Table 16: Study 1.2 Comorbid Psychiatric Diagnoses for AN

| | Full | Partial | Diagnosis[1] | Diagnostic details | | | |
				N	Severity[2]	AOO (yrs)	DOI (yrs)
Anxiety Disorders							
OCD	23 (36.5%)	8 (12.7%)	31 (49.2%)	27	3.26 (1.46)	13.82 (5.74)	8.02 (6.98)
OCPD	13 (20.0%)	8 (12.3%)	21 (32.3%)	-	-	-	-
Panic Disorder	15 (22.4%)	9 (13.4%)	24 (35.8%)	21	4.13 (1.29)	19.05 (5.13)	5.17 (5.52)
Social Phobia	19 (28.4%)	1 (1.5%)	20 (29.9%)	20	3.05 (1.32)	9.45 (5.01)	15.64 (12.77)
Specific Phobia	8 (11.9%)	4 (6.0%)	12 (17.9%)	12	2.75 (0.87)	11.25 (8.07)	11.08 (8.07)
PTSD	4 (6.0%)	1 (1.5%)	5 (7.5%)	5	3.00 (1.41)	20.40 (11.48)	10.80 (8.64)
GAD	8 (11.9%)	0	8 (11.9%)	8	2.25 (0.46)	12.38 (6.46)	12.88 (10.19)
BDD	0	1 (1.5%)	1 (1.5%)	1	3.00 (-)	43.00 (-)	1.00 (-)
Mood Disorders							
MDD	41 (61.2%)	7 (10.4%)	48 (71.6%)	42	4.30 (1.32)	16.12 (4.72)	4.21 (4.62) [3]
Bipolar	0	1 (1.5%)	1 (1.5%)	1	4.00 (-)	17.00 (-)	2.00 (-)
Dysthymia	2 (3.0%)	1 (1.5%)	3 (4.5%)	2	5.50 (0.71)	15.50 (0.71)	2.50 (0.71)
Substance Disorders							
Alcohol A/D	11 (16.4%)	2 (3.0%)	13 (19.4%)	12	4.67 (1.61)	20.42 (5.85)	1.78 (1.21)
Sub. A/D	10 (15.0%)	1 (1.5%)	11 (16.5%)	7	4.86 (1.07)	17.00 (1.91)	3.60 (1.52)

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder; Alcohol A/D Alcohol abuse/dependence; Sub. A/D Substance abuse/dependence.

[1]Diagnosis indicates pooled data from full and partial (1 criterion short) diagnosis. Severity, AOO and DOI details are for pooled data.

[2]Severity rated on 6-point scale from 1 (severe) to 6 (prior history), where 5 and 6 indicate fully recovered (>1 year)

[3]Period of time over which depressive episode/s occurred.

Table 17: Study 1.2 Comorbid Psychiatric Diagnoses for BN

	Full	Partial	Diagnosis[1]	Diagnostic details			
				N	Severity[2]	AOO (yrs)	DOI (yrs)
Anxiety Disorders							
OCD	12 (40%)	2 (6.7%)	14 (46.7%)	12	3.50 (1.24)	13.70 (4.99)	8.25 (6.60)
OCPD	2 (7.1%)	9 (32.1%)	11 (39.2%)	-	-	-	-
Panic Disorder	3 (10%)	2 (6.7%)	5 (16.7%)	5	4.00 (1.10)	29.40 (7.50)	2.72 (2.36)
Social Phobia	13 (43.3%)	6 (20%)	19 (63.3%)	19	2.21 (0.86)	11.11 (7.49)	16.81 (10.95)
Specific Phobia	5 (16.7%)	5 (16.7%)	10 (33.4%)	10	2.00 (1.05)	10.90 (6.76)	17.90 (8.28)
PTSD	4 (13.3%)	2 (6.7%)	6 (20%)	6	3.67 (1.37)	17.00 (7.43)	2.73 (3.14)
GAD	6 (20%)	3 (10%)	9 (30%)	9	2.33 (0.50)	11.89 (8.54)	15.56 (7.99)
BDD	0	0	0	0	-	-	-
Mood Disorders							
MDD	19 (63.3%)	4 (13.3%)	23 (76.6%)	21	3.57 (1.50)	17.62 (5.50)	8.32 (8.56)[3]
Bipolar	3 (10%)	1 (3.3%)	4 (13.3%)	4	3.75 (0.50)	16.75 (1.71)	5.38 (4.19)
Dysthymia	0	0	0	0	-	-	-
Substance Disorders							
Alcohol A/D	13 (43.3%)	2 (6.7%)	15 (50%)	15	3.40 (1.60)	20.33 (5.80)	5.36 (7.80)
Sub. A/D	10 (33.3%)	0	10 (33.3%)	10	4.60 (1.84)	21.50 (5.91)	5.78 (4.55)

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder; Alcohol A/D Alcohol abuse/dependence; Sub. A/D Substance abuse/dependence.

[1]Diagnosis indicates pooled data from full and partial (1 criterion short) diagnosis. Severity, AOO and DOI details are for pooled data.

[2]Severity rated on 6-point scale from 1 (severe) to 6 (prior history), where 5 and 6 indicate fully recovered (>1 year)

[3]Period of time over which depressive episode/s occurred.

Table 18: Study 1.2 Comorbidity diagnostic subtyping for OCD, substance abuse/dependence, and specific phobia in addition to self-harm prevalence.

	AN	BN
Self-harm	25 (38.5%)	16 (55%)
OCD sybtype (prominent)		
Ordering	15 (50.0%)	5 (38.5%)
Checking	1 (3.3%)	2 (15.4%)
Hording	1 (3.3%)	1 (7.7%)
Neutralising	0	1 (7.7%)
Washing	10 (33.3%)	3 (23.1%)
Obsessing	0	1 (7.7%)
Other	3 (10.0%)	0
Specific phobia subtype		
Animal	4 (33.3%)	5 (50.0%)
Natural/enviro.	0	2 (20.0%)
Blood/injection	3 (25.0%)	1 (10.0%)
Situational	4 (33.3%)	2 (20.0%)
Other	1 (8.3%)	0
Substance A/D type		
Sedatives	2 (20.0%)	0
Cannabis	4 (40.0%)	4 (40.0%)
Stimulants	0	0
Opioids	0	1 (10.0%)
Cocaine	1 (10.0%)	3 (30.0%)
Hallucinogens	0	1 (10.0%)
Other/Poly	3 (30.0%)	1 (10.0%)

AN Anorexia Nervosa; BN Bulimia Nervosa; OCD Obsessive-compulsive disorder; A/D Abuse/Dependence

endorsed ordering and washing, those with BN represented more the full spectrum of OCD subtypes. Threshold OCPD was higher in AN than BN groups however a large, number of BN participants fell just short of meeting a full OCPD diagnosis, with three rather than four items endorsed to threshold level.

In terms of other anxiety diagnoses, social phobia was surprisingly more prevalent in the BN group, where over 40% met threshold and a further 20% met partial diagnostic criteria. While age of onset was lower in the ANR groupduration of illness was nearly twice as long in the BN group however this trend did not reach significance. Specific phobia, post traumatic stress disorder (PTSD), and generalised anxiety disorder (GAD) were also notably higher in the BN sample. Specific phobia was largely animal type across clinical groups, however situational type and blood/injection type were also endorsed.

Alcohol abuse/dependence was a notable 30% higher in the BN group, who also displayed a longer duration of illness particularly with regard to alcohol disorders. Cannabis was used across groups, with Opioids and Hallucinogens present in the BN group only (note small sample sizes).

6.2.4.3 Set-shifting results

See Table 19 for descriptive statistics, test statistics, and effect size results for set-shifting variables, split by AN, BN and HC.

Trail Making Test (TMT): A significant difference between groups was found for raw shifting time (where age was a significant covariate; p=0.01) and the controlled B-A variable. Post-hoc LSD comparisons revealed the BN group only showed longer latencies on the TMT compared to HC, with moderate to large effect sizes across both outcome variables (raw and controlled). A Kruskal-Wallis (KW) test indicated no overall difference on the number of errors made in the shifting trial, however a direct comparison between BN and HC groups using a Mann-Whitney U test revealed significantly more errors in the BN group, with a moderate effect size.

Wisconsin Card Sorting Test (WCST): A significant group difference on the WCST was found, where the AN group made significantly more perseverative errors (with a large upper quartile) and completed less categories than the HC group. The BN group did not differ significantly from the HC group on perseverative errors (p=0.06) although a trend was observed in the expected direction with a small effect size. The BN group completed less categories than HC (p=0.02). Notable variance was observed within the ED groups, particularly for AN.

Table 19: Study 1.2 Set-shifting descriptive statistics by diagnostic group

	AN	BN	HC	Test Statistic			Cohen's d [2]	
	(n=62)	(n=26)	(n=78)	F	KW	p-value	AN	BN
TMT shift time (B) [3]	29.31 (7.41)	33.60 (9.02)	28.08 (6.92)	6.20	-	<0.01	0.15	0.61**
TMT B-A	9.16 (6.27)	13.69 (7.04)	8.89 (6.31)	6.50	-	<0.01	0.04	0.74**
TMT errors (shift) [1]	0 (0-1)	1 (0-1)	0 (0-1)	-	4.90	0.09	0.06	0.47*
WCST Perseverative errors [1]	9 (7-21)	9 (6-13)	7 (5.75-9)	-	11.85	<0.01	0.57**	0.33
WCST Categories completed [1]	6 (5-6)	6 (6-6)	6 (6-6)	-	24.25	<0.001	0.88**	0.48*
Brixton errors	11.72 (3.97)	10.37 (4.30)	10.01 (4.21)	3.38	-	0.04 [£]	0.42**	0.09
CatBat shift time (bat)	31.50 (10.49)	30.39 (8.25)	29.08 (11.02)	0.96	-	0.39	0.22	0.13
CatBat B-C	9.37 (6.22)	10.11 (9.80)	8.38 (7.58)	0.65	-	0.52	0.14	0.21
CatBat errors (shift) [1]	0 (0-1)	1 (0-1)	0 (0-1)	-	3.55	0.17	0.03	0.33
Haptic perseverations [1]	16 (10-30)	22 (11-30)	13 (7-21.75)	-	5.83	0.05	0.30	0.38*

AN Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control; KW Kruskal-Wallis Test; TMT Trail Making Test; WCST Wisconsin Card Sorting Test
[1]Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses
[2] Cohen's d effect size comparisons with HC data
[3]Age run as covariate
* Comparison significant at 0.05 level
** Comparison significant at 0.01 level
[£] Comparison no longer significant after Hochberg correction

Brixton Task: A significant group difference on the Brixton task was found, where the AN group made significantly more errors than HC. No difference was found between BN and HC groups, with a negligible effect size.

CatBat Task: No significant differences between groups were found for any outcome measure on the CatBat task.

Haptic Illusion: The analysis for the Haptic task just reached significance, where the BN group showed significantly more illusions than HC (small to moderate effect size). A similar small effect size fell short of significance for the AN group (p= 0.07) although a trend was observed in the hypothesised direction, showing an almost identical range to that of the BN group. The upper quartile of both ED groups was at ceiling, with a large variance in scores across AN, BN and HC groups.

6.2.4.4 Coherence results

See Table 20 for descriptive statistics, test statistics, and effect size results for coherence measures across diagnostic groups

Group Embedded Figures Test (GEFT): A significant group difference on GEFT median time and number of time out errors was found. The AN group was both faster and made less time out errors than HC, with moderate effect sizes for both variables. The BN group did not differ to HC.

Rey-Osterrieth Complex Figure (ROCF): Significant differences on ROCF copy central coherence index and both of the coherence indices (order & style) were found. Both AN and BN groups had a lower index than HC, where effect sizes were large for AN, and large to very large for BN compared to HC.

Drawing accuracy did not differ in the AN and HC groups, however the BN group scored significantly lower than both AN (p<0.01) and HC (p<0.01) on copy accuracy. This pattern persisted for recall accuracy (AN p<0.04; HC p=0.001), where age was a significant covariate (p<0.001). A significant moderate positive correlation was found between copy coherence index and recall accuracy across all groups (r(179)=0.42, p<0.001), indicating that the more coherent copiers remembered more of the drawing 20 minutes later.

Table 20: Study 1.2 Coherence descriptive statistics by diagnostic group.

	AN (n=68)	BN (n=28)	HC (n=81)	F	KW	p-value	AN	BN
				Test Statistic			Cohen's d [2]	
GEFT median [1]	6.63 (4.63-10.74)	8.53 (6.25-11.10)	8.85 (5.86-15.43)	-	7.99	0.02	-0.47**	0.08
GEFT time out errors [1]	1 (0-2)	1 (0-2)	1 (1-2.75)	-	7.62	0.02	0.46**	0.08
ROCF coherence index [1]	1.42 (1-1.61)	1.31 (0.94-1.54)	1.56 (1.41-1.56)	-	23.04	<0.001	-0.63**	0.87**
ROCF order [1]	2.17 (1.81-2.5)	1.83 (1.5-2.33)	2.45 (2.17-2.67)	-	23.22	<0.001	-0.58**	0.94**
ROCF style [1]	1.5 (1-1.67)	1.33 (0.83-1.67)	1.67 (1.5-1.83)	-	19.00	<0.001	-0.63**	0.71**
ROCF copy accuracy	29.43 (3.00)	26.85 (3.94)	29.31 (3.92)	6.11	-	<0.01	0.03	-0.63**
ROCF recall accuracy [3]	15.87 (5.56)	13.21 (5.07)	16.58 (4.87)	7.94	-	<0.001	-0.14	-0.68**

AN Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control; KW Kruskal-Wallis Test; GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure
[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses
[2] Cohen's d effect size comparisons with HC data
[3] Age run as covariate
* Comparison significant at 0.05 level
** Comparison significant at 0.01 level

6.2.5   Method Analysis 3: Lifetime phenotype split

To further explore the findings already presented in this chapter, results will be analysed in one additional orientation this time classified by lifetime clinical presentation rather than traditional diagnostic categories. Given the phenotypic overlap in the subtypes of AN and BN, both of which can be characterised by restricting, purging and binging behaviours, in many cases a traditional diagnosis of AN or BN is largely based on BMI (±17.5 BMI) rather than clinical symptomology. The current split will ignore weight as a factor, instead classifying participants by lifetime phenotypic behaviour; restricting only (ANR), bulimic behaviours at any weight to include concurrent presentation (binge/purging AN) or presentation over time (BN with a history of AN; ANBN), and bulimia only (BN) compared to HC.

6.2.5.1   Participants

Participants for this study were those previously reported in analyses 1 and 2, split by lifetime phenotypic behaviour; ANR (n=35), ANBN (n=47; 24 ANP, 9 ANBP, 13 with history of BN), and BN (n=17) compared to the HC group (n=88).

6.2.5.2   Statistical methods

Normality assessment across the four groups again revealed the same pattern of distribution across both neuropsychological and self-report measures as that previously reported in the transdiagnostic split. Analyses will be the same as that outlined in method 2 (6.2.3.2).

6.2.6   Results Analysis 3: Lifetime Phenotype split

6.2.6.1   Demographic and Clinical Features

The ANR group was significantly younger and had less years of education than the ANBN, BN and HC groups who were comparable (see Table 21). Both the ANR and ANBN groups had a significantly lower current BMI than BN and HC groups, with the lowest ever BMI following the same pattern. Clinical groups did not differ on current severity or age of onset, however the ANR group had a significantly shorter duration of illness compared to both ANBN and BN groups. YBC-EDS scores did not differ significantly, although the ANBN group showed a trend toward a higher score, indicating a more severe illness at its worst. Large variation was seen within all clinical groups on the Y-BOCS.

Self-report Clinical Features: Significant differences were found between all three clinical groups and the HC group across all measures (see Table 22), except the

Table 21: Study 1.3 Demographic and clinical features

	ANR	ANBN	BN	HC	Test statistic	
	(n=35)	(n=46)	(n=17)	(n=88)	F	p-value
Age [a]	23.71 (6.39)	26.11 (7.60)	25.65 (6.29)	28.43 (8.47)	3.41	0.02*
Years of Education [a]	15.37 (2.70)	15.90 (4.24)	15.61 (3.31)	16.76 (1.98)	3.27	0.02*
BMI (current) [a, b]	17.98 (2.18)	18.58 (3.23)	22.66 (2.16)	22.07 (1.79)	41.46	<0.001**
BMI (lowest) [c, f]	14.25 (1.84)	14.92 (5.03)	19.62 (1.67)	-	13.15	<0.001**
BMI (highest) [c, f]	21.38 (2.94)	22.60 (3.14)	25.71 (3.36)	-	11.04	<0.001**
Current Severity	2.60 (1.04)	2.30 (1.13)	2.35 (1.12)	-	0.76	0.47
Age of ED Onset	17.20 (4.65)	15.94 (3.45)	18.18 (3.86)	-	2.28	0.11
Duration of Illness [d]	6.34 (4.57)	9.90 (6.55)	7.09 (5.94)	-	4.07	0.02*

ANR Restricting type Anorexia Nervosa; ANBN Mixed AN and BN behaviours; BN Bulimia Nervosa; HC Healthy Control; BMI Body mass index
[a] Significant difference (Tukey's HSD) between ANR and HC
[b] Significant difference (Tukey's HSD) between ANBP/ANBN and HC
[c] Significant difference (Tukey's HSD) between ANR and BN
[d] Significant difference (Tukey's HSD) between ANR and ANBP/ANBN
[f] Significant difference (Tukey's HSD) between ANBP/ANBN and BN
* Comparison significant at 0.05 level
** Comparison significant at 0.01 level

Table 22: Study 1.3 Self-report clinical features

	ANR	ANBN	BN	HC	Test statistic			Cohen's d [2]		
	(n=35)	(n=43)	(n=17)	(n=88)	F	KW	p-value	ANR	ANBN	BN
HADS anxiety	10.59 (5.28)	12.76 (4.03)	13.53 (4.50)	4.20 (2.32)	72.52	-	<0.001	1.87**	2.87**	3.36**
HADS depression [1]	5 (3-7)	8 (4-12)	7 (3-11.5)	1 (0-2)	-	100.98	<0.001	1.61**	2.17**	1.48**
OCI-R total [1]	16 (4-28)	18 (12.38-29)	22 (11.5-35.5)	5.5 (2.25-10)	-	64.48	<0.001	0.86**	1.54**	1.17**
Rosenberg self-esteem	13.80 (5.79)	9.37 (5.31)	12.88 (4.48)	23.51 (3.96)	101.32	-	<0.001	-2.13**	-3.18**	-2.63**
Frost Perfectionism	93.25 (16.33)	100.26 (16.20)	94.09 (13.40)	72.95 (15.22) [3]	14.44	-	<0.001	1.31**	-4.94**	1.41**
CHiRP total [1]	3 (1-4)	3 (1-4)	1 (0-3)	1 (0-2) [3]	-	22.52	<0.001	1.39**	1.44**	0.67
Y-BOCS	14.56 (13.66)	13.38 (12.56)	12.53 (11.64)	-	0.16	-	0.85	-		
YBC-EDS	23.88 (5.45)	25.05 (4.13)	21.24 (7.88)	-	2.96	-	0.06	-		
Thinking Styles	23.26 (6.69)	25.47 (6.79)	29.54 (7.75)	14.66 (4.89)	52.35	-	<0.001	1.58**	1.92**	2.77**
Cognitive Flexibility	50.05 (8.57)	46.09 (8.93)	44.84 (11.86)	60.34 (5.96)	42.97	-	<0.001	-1.52**	-1.99**	-2.20**

ANR Restricting type Anorexia Nervosa; ANBN Mixed AN and BN behaviours; BN Bulimia Nervosa; HC Healthy Control; KW Kruskal-Wallis Test; HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire; YBC-EDS Yale-Brown-Cornell Eating Disorder Scale; Y-BOCS Yale-Brown Obsessive-Compulsive Scale.

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons with HC data

[3] HC data collected from a subset of participants (n=22)

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

CHiRP where BN did not differ from HC. ANR reported significantly lower depression levels and higher self-esteem than the mixed ANBN group, where pure BN fell between the two.

Self-report Cognitive Style: Significant differences were found between all clinical subtypes and HC on the CFS and TSQ. Very large effect sizes were seen.

6.2.6.2 Comorbidity

Details of comorbid diagnoses across groups can be found in Table 23, Table 24, and Table 25. Comorbidity across the three clinical groups showed consistently high rates of depression, with over 80% of those from the mixed ANBN cohort reporting lifetime MDD with an average 7-year illness duration. Just over 70% of the BN group also presented with MDD with a similar illness duration. Rates for ANR were slightly reduced at just over 60%, however a notably shorter illness duration of under three years was reported in this group (Cohen's d = 0.95 compared to ANBN). Self-harm rates were highest in the ANBN group at 64.3%, with just over half of the BN sample also reporting self-harm (52.9%) but a low rate of 14.3% in ANR.

OCD was the most endorsed anxiety disorder in ANR and ANBN groups, however despite a similar OCD prevalence rate in the BN group, social phobia was most prevalent amongst pure BN with 70% (or 12 of 17 women) meeting threshold or subthreshold criteria. Panic disorder in the ANBN group was present for on average two or three times the duration as in ANR and pure BN groups. Age of onset for GAD was substantially later in the ANBN group at approximately 17 years of age in comparison to ANR and pure BN groups where 7.5 and 8.2 years (respectively) were reported. One case of ANR with subthreshold BDD was present. Overall anxiety disorders were least common in the ANR group, where 51.4% of cases reported no lifetime anxiety disorder (threshold or subthreshold), compared to only 20% in the ANBN group and 11.8% (2 women) in the pure BN group.

The presence of an alcohol or substance disorder was predictably highest in the BN group, with over half reporting alcohol abuse/dependence and nearly 30% reporting substance abuse/dependence. This dropped to 27% and 29% (respectively) in the ANBN group, and further to 20% and 8.6% in the ANR group where only 1 individual met criteria for threshold substance disorder.

6.2.6.3 Set-shifting results

See for descriptive statistics, test statistics and effect size results for set-shifting tasks across ANR, ANBN, BN and HC groups.

Table 23: Study 1.3 Comorbid Psychiatric Diagnoses for ANR

	Full	Partial	Diagnosis[1]	Diagnostic details			
				N	Severity[2]	AOO (yrs)	DOI (yrs)
Anxiety Disorders							
OCD	13 (39.4%)	4 (12.1%)	17 (51.5%)	15	3.20 (1.61)	13.20 (4.78)	7.68 (6.92)
OCPD	7 (20%)	4 (11.4%)	11 (31.4%)	-	-	-	-
Panic Disorder	6 (17.1%)	3 (8.6%)	9 (25.7%)	9	4.56 (1.51)	17.38 (1.92)	3.03 (4.98)
Social Phobia	8 (22.9%)	0	8 (22.9%)	8	2.38 (0.74)	7.38 (2.56)	9.83 (5.08)
Specific Phobia	4 (11.4%)	2 (5.7%)	6 (17.1%)	6	2.67 (1.21)	11.83 (6.46)	9.83 (5.08)
PTSD	1 (2.9%)	0	1 (2.9%)	1	3.00 (-)	39.00 (-)	5.0 (-)
GAD	4 (11.4%)	0	4 (11.4%)	4	2.25 (0.50)	7.5 (3.00)	20.75 (7.89)
BDD	0	1 (2.9%)	1 (2.9%)	1	3.00 (-)	43.00 (-)	1.00 (-)
Mood Disorders							
MDD	19 (54.3%)	3 (8.6%)	22 (62.9%)	21	4.86 (0.96)	15.65 (4.32)	2.57 (3.11)[3]
Bipolar	0	0	0	0	-	-	-
Dysthymia	1 (2.9%)	1 (2.9%)	2 (5.4%)	2	5.50 (0.71)	15.50 (0.71)	2.50 (0.71)
Substance Disorders							
Alcohol A/D	5 (14.3%)	2 (5.7%)	7 (20%)	6	5.33 (0.82)	17.67 (1.21)	1.3 (0.97)
Sub. A/D	1 (2.9%)	2 (5.7%)	3 (8.6%)	2	5.50 (0.71)	14.50 (0.71)	5.00 (-)

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder; Alcohol A/D Alcohol abuse/dependence; Sub. A/D Substance abuse/dependence.

[1]Diagnosis indicates pooled data from full and partial (1 criterion short) diagnosis. Severity, AOO and DOI details are for pooled data.

[2]Severity rated on 6-point scale from 1 (severe) to 6 (prior history), where 5 and 6 indicate fully recovered (>1 year)

[3]Period of time over which depressive episode/s occurred.

Table 24: Study 1.3 Comorbid Psychiatric Diagnoses for ANBN

| | Full | Partial | Diagnosis[1] | Diagnostic details | | | |
				N	Severity[2]	AOO (yrs)	DOI (yrs)
Anxiety Disorders							
OCD	16 (37.2%)	4 (9.3%)	20 (43.5%)	17	3.47 (1.38)	13.94 (6.37)	8.25 (7.15)
OCPD	8 (18.2%)	9 (20.5%)	17 (38.7%)				
Panic Disorder	9 (20.0%)	6 (13.3%)	15 (33.3%)	15	3.87 (1.06)	21.29 (7.50)	6.54 (5.36)
Social Phobia	18 (40.0%)	1 (2.2%)	19 (42.2%)	19	2.84 (1.46)	9.79 (5.66)	16.95 (13.31)
Specific Phobia	8 (17.8%)	4 (8.9%)	12 (26.7%)	12	2.50 (0.91)	11.25 (8.66)	16.58 (8.43)
PTSD	4 (8.9%)	3 (6.7%)	7 (15.6%)	7	3.71 (1.50)	18.29 (5.53)	7.49 (8.85)
GAD	6 (13.3%)	2 (4.4%)	8 (17.7%)	8	2.38 (0.52)	16.88 (7.64)	8.88 (6.22)
BDD	0	0	0	-			
Mood Disorders							
MDD	32 (71.1%)	5 (11.1%)	37 (82.2%)	33	3.61 (1.39)	16.61 (4.98)	7.10 (7.10)[3]
Bipolar	1 (2.2%)	2 (4.4%)	3 (6.6%)	3	4.00 (-)	17.67 (1.15)	3.67 (2.08)
Dysthymia	1 (2.2%)	0	1 (2.2%)	0	-		
Substance Disorders							
Alcohol A/D	12 (26.7%)	0	12 (26.7%)	12	3.83 (1.70)	21.92 (6.22)	4.22 (7.65)
Sub. A/D	13 (28.9%)	0	13 (28.9%)	10	4.70 (1.34)	20.70 (6.09)	4.00 (3.85)

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder; Alcohol A/D Alcohol abuse/dependence; Sub. A/D Substance abuse/dependence.

[1]Diagnosis indicates pooled data from full and partial (1 criterion short) diagnosis. Severity, AOO and DOI details are for pooled data.

[2]Severity rated on 6-point scale from 1 (severe) to 6 (prior history), where 5 and 6 indicate fully recovered (>1 year)

[3]Period of time over which depressive episode/s occurred.

Table 25: Study 1.3 Comorbid Psychiatric Diagnoses for BN

	Full	Partial	Diagnosis[1]	Diagnostic details			
				N	Severity[2]	AOO (yrs)	DOI (yrs)
Anxiety Disorders							
OCD	6 (35.3%)	2 (11.8%)	8 (47.1%)	7	3.29 (0.95)	14.83 (5.27)	8.67 (6.59)
OCPD	0	4 (25.0%)	4 (25.0%)	-			
Panic Disorder	4 (23.5%)	1 (5.9%)	5 (29.4%)	5	4.00 (1.23)	27.50 (7.14)	2.15 (2.29)
Social Phobia	6 (35.3%)	6 (35.3%)	12 (70.6%)	12	2.50 (0.91)	12.92 (8.17)	13.36 (9.91)
Specific Phobia	1 (5.9%)	3 (17.6%)	4 (23.5%)	4	1.75 (0.96)	9.50 (5.20)	13.50 (5.74)
PTSD	3 (17.6%)	0	3 (17.6%)	3	2.67 (1.16)	12.33 (7.64)	4.33 (4.04)
GAD	4 (23.5%)	1 (5.9%)	5 (29.4%)	5	2.20 (0.45)	8.20 (4.92)	17.80 (9.18)
BDD	0	0	0	-			
Mood Disorders							
MDD	9 (52.9%)	3 (17.6%)	12 (70.5%)	10	3.90 (1.66)	18.60 (6.19)	6.60 (7.52)[3]
Bipolar	2 (11.8%)	0	2 (11.8%)	2	3.50 (0.71)	15.50 (0.71)	6.25 (6.72)
Dysthymia	0	0	0	-			
Substance Disorders							
Alcohol A/D	7 (41.2%)	2 (11.8%)	9 (53.0%)	9	3.22 (1.72)	20.11 (6.55)	4.69 (5.61)
Sub. A/D	5 (29.4%)	0	5 (29.4%)	5	4.40 (2.19)	19.60 (2.30)	6.60 (4.04)

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder; Alcohol A/D Alcohol abuse/dependence; Sub. A/D Substance abuse/dependence.

[1]Diagnosis indicates pooled data from full and partial (1 criterion short) diagnosis. Severity, AOO and DOI details are for pooled data.

[2]Severity rated on 6-point scale from 1 (severe) to 6 (prior history), where 5 and 6 indicate fully recovered (>1 year)

[3]Period of time over which depressive episode/s occurred.

Table 26: Study 1.3 Set-shifting descriptive statistics by lifetime phenotype split

	ANR	ANBN	BN	HC	Test Statistic			Cohen's d [2]		
	(n=31)	(n=43)	(n=13)	(n=78)	F	KW	p-value	ANR	ANBN	BN
TMT shift time (B) [3]	28.97 (7.28)	30.98 (8.53)	33.10 (8.42)	28.08 (6.92)	3.82	-	<0.01	0.13	0.39*	0.70**
TMT B-A	8.82 (5.71)	10.63 (7.25)	13.73 (6.81)	8.89 (6.31)	3.11	-	0.03 [£]	-0.01	0.26	0.76**
TMT shift errors (B) [1]	0 (0-1)	0 (0-1)	1 (0-1)	0 (0-1)	-	4.45	0.22	0.24	0.29	0.34
WCST Perseverative errors [1]	8 (6-15.5)	9 (7-21.75)	11 (6.5-16.5)	7 (5.75-9)	-	11.54	<0.01	0.30	0.63**	0.46*
WCST Categories completed [1]	6 (5.5-6)	6 (5-6)	6 (6-6)	6 (6-6)	-	20.21	<0.001	0.84**	0.87**	0.58**
Brixton errors	10.66 (4.00)	11.91 (4.37)	11.00 (3.48)	10.01 (4.21)	2.14	-	0.10	0.16	0.45**	0.24
CatBat shift time (B)	31.05 (10.88)	31.98 (9.71)	29.16 (8.07)	29.08 (11.02)	0.85	-	0.47	0.18	0.27	0.01
CatBat B-C	8.92 (6.43)	10.49 (7.03)	8.51 (10.26)	8.38 (7.58)	0.77	-	0.51	0.07	0.29	0.02
CatBat shift errors (B) [1]	0 (0-1)	0 (0-1)	1 (0-2)	0 (0-1)	-	4.54	0.21	-0.01	0.01	0.42*
Haptic perseverations [1]	18 (11-30)	15.5 (9-30)	16 (9.5-30)	13 (7-21.75)	-	6.04	0.11	0.42*	0.27	0.17

AN Anorexia Nervosa; ANBN Mixed AN and BN behaviours; BN Bulimia Nervosa; HC Healthy Control; KW Kruskal-Wallis Test; TMT Trail Making Test; WCST Wisconsin Card Sorting Test

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons with HC data

[3] Age run as covariate

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

[£] Comparison no longer significant after Hochberg correction

TMT: A significant group difference was found for raw TMT time, where age was again a significant covariate (p=0.01). Both the ANBN and pure BN groups showed longer shift times, which was most notable in the pure BN group (large effect size). This difference persisted for the pure BN group only on the controlled B-A variable. The ANR group was comparable to HC across variables. No significant differences were found for number of errors made.

WCST: A significant group difference for perseverative errors and categories completed on the WCST was found, where the mixed ANBN group showed the most prominent difficulties. Perseverative errors in the ANR group did not differ significantly from HC (p=0.12) despite a notably higher upper quartile, suggesting larger variance in this group. All clinical groups completed significantly less categories than the HC group, despite the lower quartile for BN remaining at ceiling. ANR and ANBN groups showed a large effect size.

Brixton Task: The overall group difference for the Brixton task did not reach significance, however post-hoc comparisons revealed a significant difference between ANBN and HC groups (p<0.01) with a moderate effect size.

CatBat Task: No overall group analyses reached significance for CatBat variables. One significant post-hoc test was found between BN and HC groups on the number of perseverative errors, where the BN group made more errors than HC. This difference reached a moderate effect size.

Haptic Illusion: Again, despite no overall group differences, post-hoc analyses revealed the ANR group made significantly more perseverations than HC. The ANBN group showed a trend in the same direction but did not reach significance (p=0.12), nor did the BN group (p=0.39). All clinical groups had upper quartiles at ceiling (30) indicating large variation within the clinical groups.

6.2.6.4   Coherence results

See Table 27 for descriptive statistics, test statistics and effect size results for coherence tasks split by ANR, ANBN, BN and HC groups

GEFT: Significant group differences were found across GEFT variables. The ANR group were both significantly faster at finding shapes (moderate effect size) and made significantly less time-out errors than HC. The ANBN group also made significantly less time-out errors and showed a trend toward a faster median time (small effect size), however this difference fell short of significance (p=0.07). The

Table 27: Study 1.3 Coherence descriptive statistics by phenotype group

	ANR	ANBN	BN	HC	Test Statistic			Cohen's d [2]		
	(n=32)	(n=44)	(n=15)	(n=81)	F	KW	p-value	ANR	ANBN	BN
GEFT median [1]	6.35 (4.63-8.99)	7.93 (5.13-11.26)	8.1 (5.6-19.95)	8.85 (5.86-15.43)	-	8.31	0.04	0.48*	0.33	0.02
GEFT time out fails [1]	1 (0-2)	1 (0-2)	2 (1-4)	1 (1-2.75)	-	11.22	0.01	0.40*	0.47*	0.16
ROCF coherence index [1]	1.43 (0.95-1.61)	1.42 (1.13-1.65)	1.24 (0.92-1.44)	1.56 (1.41-1.68)	-	23.04	<0.001	0.61**	0.54**	1.00**
ROCF order [1]	2.17 (1.83-2.5)	2.17 (1.68-2.5)	1.83 (1.5-2.09)	2.45 (2.17-2.67)	-	23.12	<0.001	0.45*	0.63**	1.04**
ROCF style [1]	1.5 (1-1.67)	1.5 (0.96-1.7)	1.33 (0.83-1.59)	1.67 (1.5-1.83)	-	19.00	<0.001	0.72**	0.45**	0.78**
ROCF copy accuracy	29.74 (2.54)	28.82 (3.80)	25.91 (3.09)	29.31 (3.92)	4.93	-	<0.01	0.12	-0.13	-0.90**
ROCF recall accuracy [3]	16.19 (5.93)	15.42 (5.18)	11.62 (4.39)	16.58 (4.87)	7.22	-	<0.001	-0.08	-0.23	-1.04**

ANR Restricting type Anorexia Nervosa; ANBN Mixed AN and BN behaviours; BN Bulimia Nervosa; HC Healthy Control; KW Kruskal-Wallis Test; GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure
[1]Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses
[2] Cohen's d effect size comparisons with HC data
[3]Age run as covariate, therefore F statistic presented
* Comparison significant at 0.05 level
** Comparison significant at 0.01 level

BN group did not differ compared to HC on either variable; with the upper quartile of BN median times 4.5 seconds slower than that of HC.

ROCF: Low scores on the central coherence index were seen across all clinical groups compared to HC, with significant group differences across both the central coherence total and its indices. Large to very large effect sizes are seen for ANR, ANBN and BN groups compared to HC. The effect was strongest in the BN group, who also displayed significantly poorer copy accuracy compared to HC with a large effect size. This low accuracy score persisted at recall.

6.2.7 Method Analysis 4: Extreme scores

6.2.7.1 Participants

Participants were those reported in analysis 1. For this analysis, the 98 women with ED were analysed by DSM-IV diagnostic categories; ANR (n=35), ANBP (n=33) and BN (n=30; 13 or 43.3% with a lifetime history of AN).

6.2.7.2 Statistical methods

A cut-off score for each neuropsychological task was calculated by adding or subtracting one standard deviation from the HC mean (depending on the direction of each task) for normally distributed data, or by using the 15th or 85th percentile (depending on the direction of each task) for non-normal data. Normality for cut-off scores was assessed for each task using data from the HC group. All variables except the GEFT were determined normal. The CatBat task was not included in this analysis given its lack of sensitivity in the current ED sample. Cut-off data points by task were as follows: TMT B-A ≥ 15.2; WCST perseverative errors ≥ 11; Brixton task ≥ 15; Haptic task ≥ 25; GEFT median ≥ 4.6, ROCF coherence index ≥ 1.37. Each case (clinical and control) was assessed using the cut-off score for each task, with those identified as above/below the cut-off considered to have an 'extreme score'.

Further set-shifting analyses: For set-shifting results, a composite score was created where participants were categorised as having 'impaired' set-shifting if they had an extreme score on two or more set-shifting tasks, or 'intact' shifting if they had one or no extreme scores. Data was analysed by investigating frequencies and by running Pearson's chi-square tests to investigate differences in the number of cases with extreme scores across diagnoses. Pearson's chi-square test was also employed to explore differences in comorbidity based on shifting ability. Independent-samples t-tests were used to investigate differences in demographic and clinical features

across those with impaired/intact shifting. Mann-Whitney U tests were employed for non-normal data. For categorical data, Pearson's chi-square test was used.

Further coherence analyses: It was not possible to compute the same composite variable for coherence tasks, as coherence is not a linear concept. A composite score was created by splitting data from both tasks into quartiles based on HC results. Participants were then categorised across both tasks according to their quartile placement (lower 2 quartiles [LQ] or upper 2 quartiles [UQ] on each task) to create a dimensional variable. Quadrant labels were created depending on whether the strategy employed for each task was optimal or not (see Figure 10). Those falling into the adaptive or persistent detail focus dimensions were further compared across demographic, clinical features and comorbidity.

6.2.8   Results Analysis 4: Extreme scores

6.2.8.1   Set-shifting results

Frequency data:[4] Figure 11 illustrates the percentage of participants by diagnostic group with extreme scores on each neuropsychological task. The WCST displayed the highest percentage of extreme scores across diagnostic groups (32.1% - 48.4% of those with current ED), followed by the Haptic task. Figure 12 outlines the percentage of each diagnostic group with impaired/intact set-shifting, in that extreme scores on two or more set-shifting tasks were present. Overall, 36.7% of women with current ED showed impaired shifting. The ANBP group had the highest rate of impaired shifting at nearly half (48.5%), followed by current BN (36.7%) and ANR (25.7%). Weighted percentage (by sample size) of those with impaired shifting again highlights the ANBP group as having the highest proportion of impaired shifting cases, followed by current BN and ANR (see Figure 13). The HC group represents only a small proportion of impaired cases (11.4%), indicating that poor set-shifting is relatively uncommon in the general population.

A significant difference was found between the frequency of individuals with impaired shifting across diagnostic categories, $\chi^2(3)=20.76$, p<0.001. Post-hoc chi-square tests revealed significant group differences between HC and ANR (p<0.05), ANBP (p<0.001), and BN (p<0.01), where all current ED groups had a significantly higher proportion of individuals with impaired set-shifting.

---

[4] Frequency analysis was also run on data including outliers. Results were approximately equivalent.

Figure 10: Illustration of the dimensional categorisation of performance on coherence tasks based on quartile scores

Good GEFT (LQ)

Adaptive detail focus   Persistent detail focus

Good ROCF (UQ)   Poor ROCF (LQ)

Persistent global focus   Maladaptive detail focus

Poor GEFT (UQ)
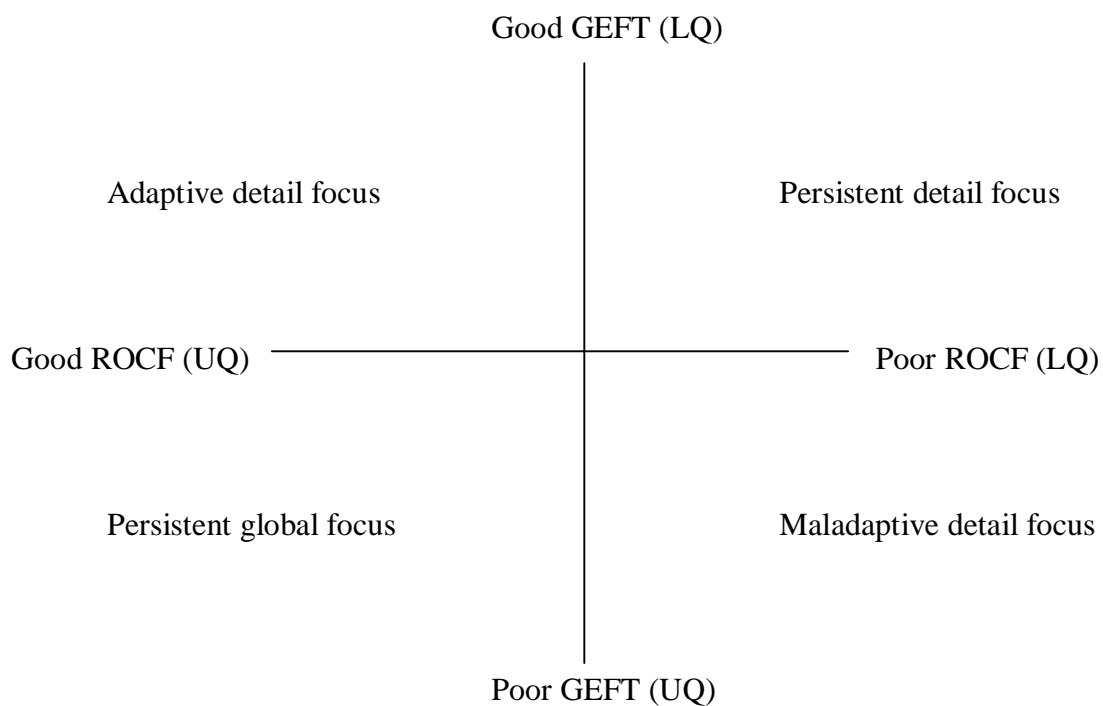
GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure; LQ Lower Quartile; UQ Upper Quartile

Figure 11: Study 1.4 Percentage of participants by diagnostic group with extreme scores on set-shifting tasks
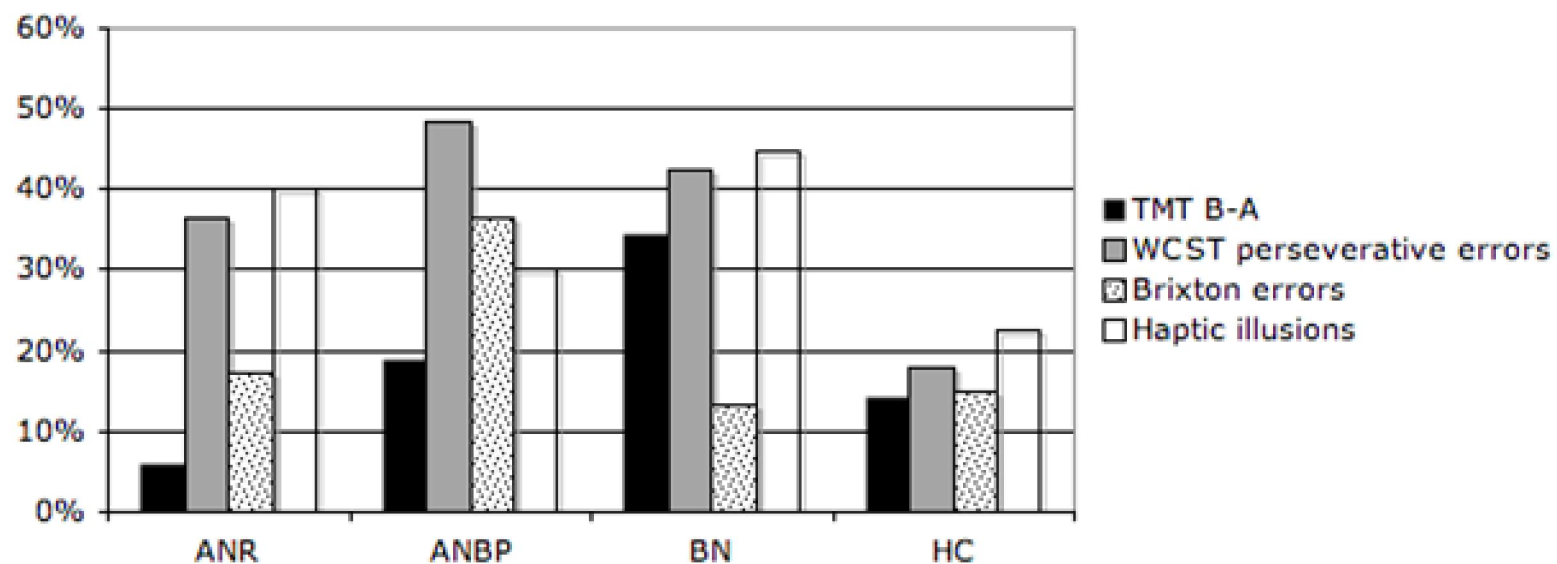
TMT Trail Making Test; WCST Wisconsin Card Sorting Test; ANR Restricting type Anorexia Nervosa; ANBP Binge/purging type Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control

Figure 12: Study 1.4 Percentage of participants by diagnostic group with impaired set-shifting ability (extreme scores on two or more tasks)
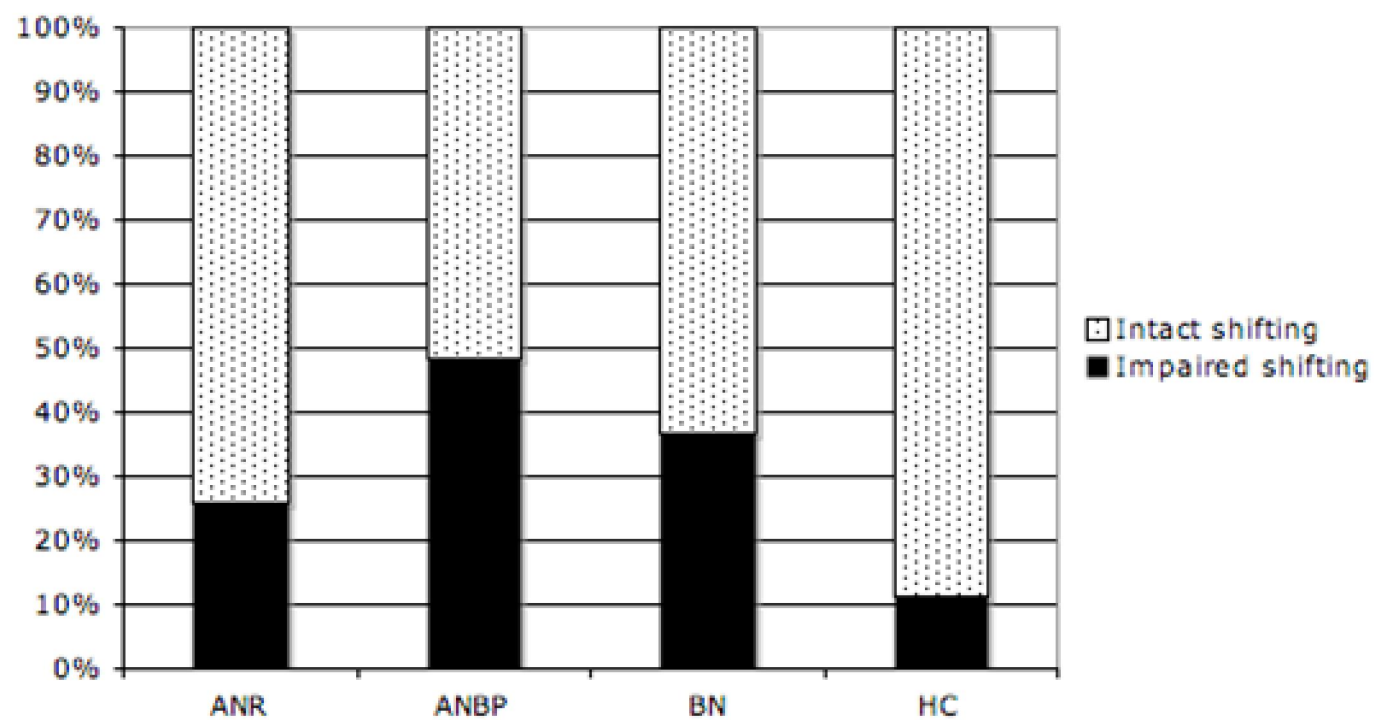
ANR Restricting type Anorexia Nervosa; ANBP Binge/purging type Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control

Figure 13: Study 1.4 Weighted percentage of participants by diagnostic group with impaired set-shifting ability
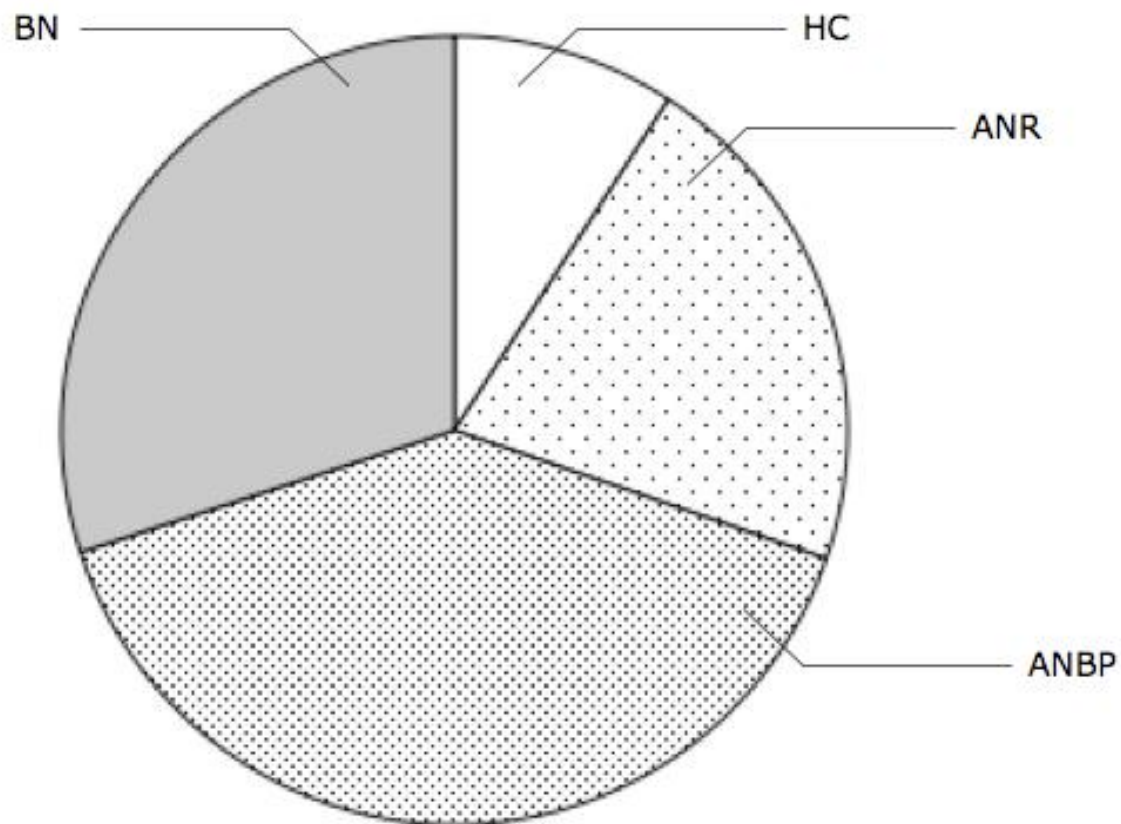
ANR Restricting type Anorexia Nervosa; ANBP Binge/purging type Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control

Demographic and clinical features:[5] Significant differences were found between those with impaired and intact set-shifting on duration of illness and YBC-EDS rituals (see Table 28). Those with impaired shifting had a longer duration of illness with a moderate effect size (see Figure 14), and more severe eating related rituals at the worst stage of their illness with a large effect size (see Figure 15). No effects of education level, weight, or current severity of the ED were found. Wide variation is noted for lowest ever BMI. While current medication did not differ significantly between groups, the impaired shifting group showed a trend toward higher medication rates, with a moderate effect size.

Self-report clinical features: Groups were closely matched on self-report measures, with most comparisons revealing negligible to small effect sizes (see Table 29). A significant finding was found on the Rosenberg self-esteem scale. Those with impaired shifting had lower self-esteem compared to those with intact shifting, with a moderate effect size (see Figure 16). Women with impaired shifting also showed a trend toward higher levels of current depression, anxiety, ordering OCD, and concern over mistakes (perfectionism subscale), all with small effect sizes. Likewise, lower levels of personal standards and childhood obsessive-compulsive behaviours including inflexibility were seen in the impaired shifting group, with a small effect size.

Comorbidity: No significant group differences emerged across individual comorbid diagnoses (see Table 30). Notably higher levels of most anxiety disorders (panic disorder, social phobia, specific phobia, PTSD) were seen in the impaired shifting group. When compared based on the frequency of anxiety diagnoses, those with impaired shifting had significantly more comorbid anxiety diagnoses (median 2, quartiles 0.25-2) compared to those with intact shifting (median 1, quartiles 0-2), with an odd's ratio just under 3. Figure 17 and Figure 18 depict the frequency of anxiety diagnoses met across women with impaired and intact shifting.

Depression rates were high across both groups, with minimal numbers of bipolar and dysthymia diagnoses. Again the impaired shifting group showed a slightly higher proportion of women with a lifetime depression diagnosis. Self-harming behaviours were significantly higher in women with impaired shifting with

---

[5] Correlations were also run between demographic and clinical features, and original neuropsychological scores (TMT B-A; WCST perseverative errors; Brixton task; Haptic task). No correlations were significant.

Table 28: Study 1.4 Descriptive statistics for demographic and clinical variables by set-shifting ability

	Intact Shifting	Impaired Shifting	Test statistics			
	(n=62)	(n=36)	t	MW	p	Cohen's d [2]
Age	24.31 (6.05)	26.67 (8.25)	-1.63	-	0.11	0.34
Years of Education [1]	15.76 (2.53)	15.50 (2.88)	0.46	-	0.65	-0.10
Current medication	0 (0-1)	1 (0-1)	-	241.5	0.10	0.45
Current ED severity	2.42 (1.05)	2.42 (1.18)	0.01	-	0.99	0.00
Age of onset	17.23 (4.41)	16.00 (3.21)	1.46	-	0.15	-0.31
Duration of illness [eq]	7.05 (4.81)	10.03 (7.30)	-2.19	-	0.03 [£]	0.51*
YBC preoccupations [eq]	12.32 (2.62)	12.91 (1.96)	-1.22	-	-0.27	0.25
YBC rituals	10.75 (3.52)	13.06 (2.34)	-3.77	-	<0.001	0.74**
	(n=43)	(n=25)				
Current BMI	17.71 (2.45)	18.31 (2.83)	-0.93	-	0.36	0.23
Lowest BMI	14.24 (1.93)	15.10 (6.62)	-0.80	-	0.43	0.20

MW Mann-Whitney U Test; ED Eating Disorder; YBC Yale-Brown-Cornell Eating Disorder Scale; BMI Body Mass Index

[eq] Equal variances not assumed

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

[£] Comparison no longer significant after Hochberg correction

Figure 14: Study 1.4 Current ED duration of illness by set-shifting ability
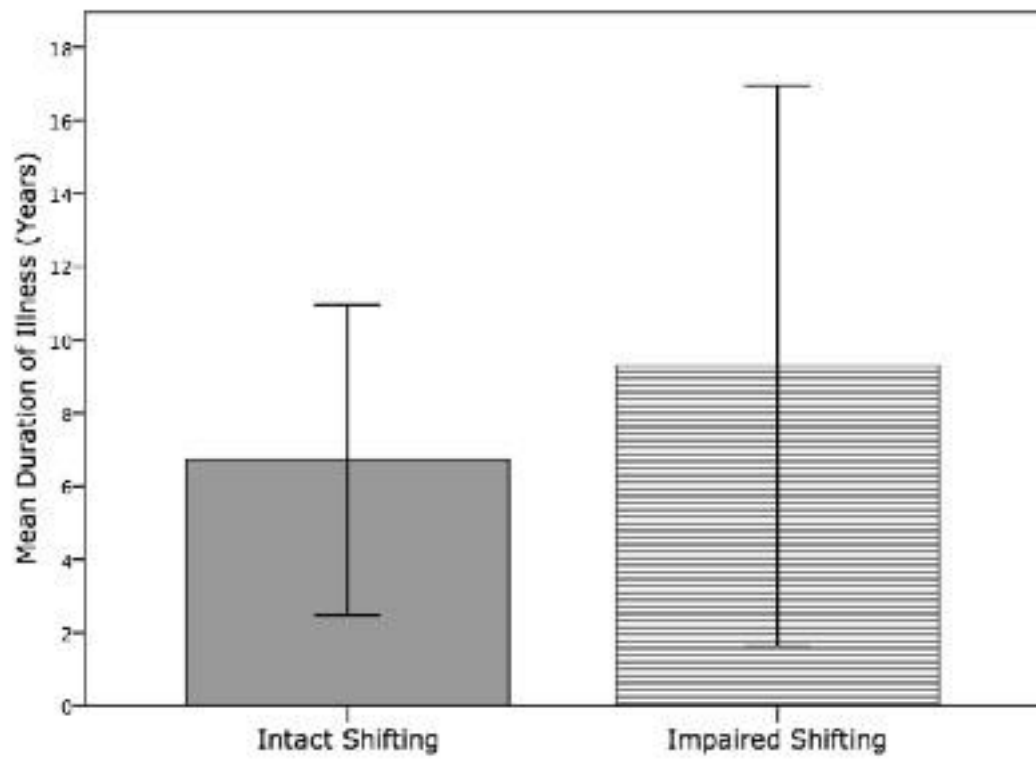
Figure 15: Study 1.4 Current ED YBC-EDS rituals by set-shifting ability
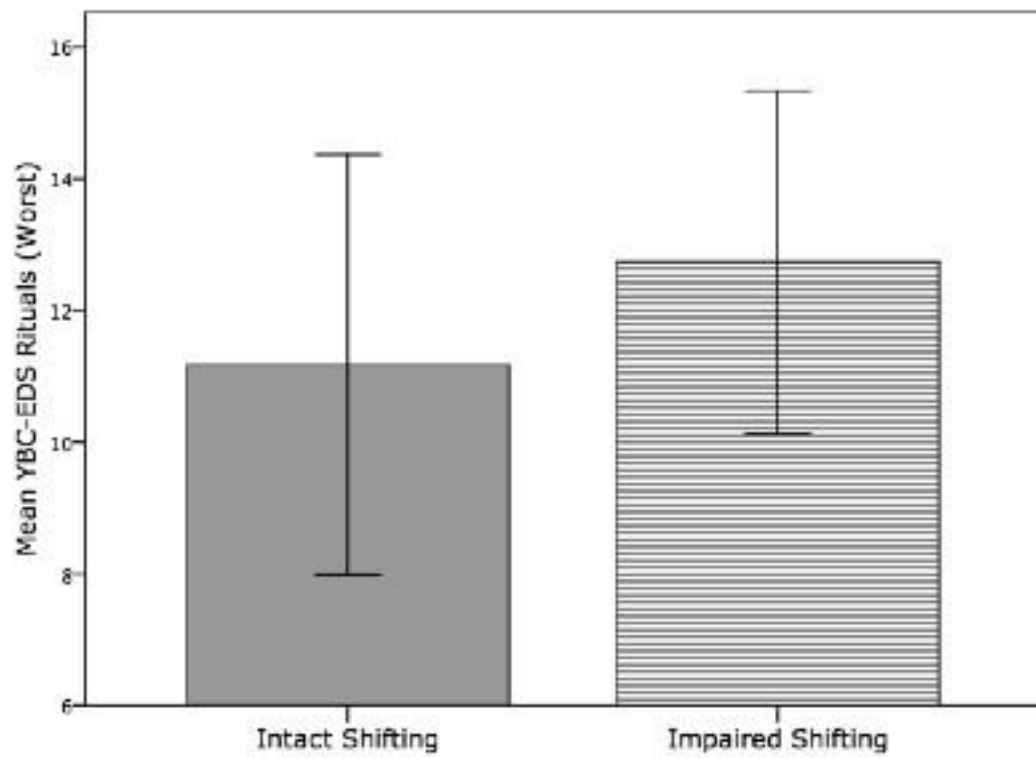
Table 29: Study 1.4 Self-report comparisons between women with current ED with intact and impaired shifting

	Intact Shifting (n=60)	Impaired Shifting (n=35)	Test statistics			
			t	MW	p	Cohen's d
HADS anxiety	11.70 (4.84)	12.74 (4.49)	-1.03	-	0.30	0.22
HADS depression	6.56 (4.36)	7.70 (4.66)	-1.20	-	0.24	0.26
OCI-R total	19.93 (14.05)	20.75 (12.41)	-0.29	-	0.77	0.06
Hording	3.68 (3.32)	3.65 (2.75)	0.04	-	0.97	-0.01
Checking	2.42 (2.94)	2.35 (1.98)	0.11	-	0.91	-0.03
Ordering	3.77 (3.36)	4.78 (3.48)	-1.39	-	0.17	0.30
Neutralising	2.30 (2.99)	2.00 (2.19)	0.51	-	0.61	-0.11
Washing	2.31 (3.29)	2.12 (2.68)	0.29	-	0.77	-0.06
Obsessing	5.46 (4.00)	5.85 (3.92)	-0.46	-	0.64	0.10
Rosenberg self-esteem	12.63 (5.71)	9.91 (5.36)	2.29	-	0.02 £	-0.49*
FMPS total	95.92 (14.74)	97.70 (18.11)	-0.52	-	0.60	0.11
Concern mistakes	31.38 (8.33)	34.11 (8.96)	-1.50	-	0.14	0.32
Personal standards	27.42 (4.42)	26.04 (5.61)	1.33	-	0.19	-0.28
Doubting	13.90 (3.36)	14.33 (3.24)	-0.62	-	0.54	0.13
Organisation	23.21 (5.96)	23.21 (5.44)	0.00	-	0.99	0.00
ChiRP total [1]	8 (4-11)	5.5 (3-11.5)	-	516.0	0.25	-0.28

Perfectionism [1]	3 (1-4)	2 (0-4)	-	690.5	0.59	-0.12
Inflexibility [1]	2 (2-3)	2 (1-3)	-	699.0	0.37	-0.20
Order/symmetry [1]	1 (0-3)	1 (0-4.5)	-	910.5	0.91	0.03

MW Mann-Whitney U Test; HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire

[1]Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

* Comparison significant at 0.05 level

£ Comparison no longer significant after Hochberg correction

Figure 16: Study 1.4 Self-esteem score across current ED with impaired and intact set-shifting
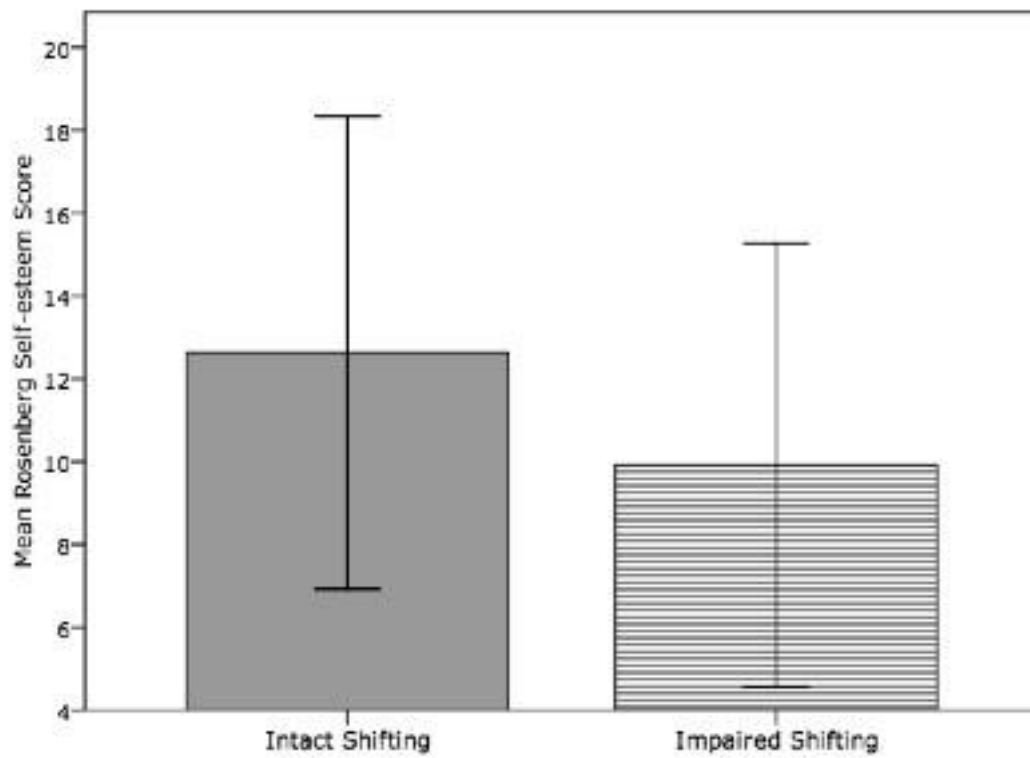
Table 30: Study 1.4 Diagnostic comparisons between current ED with intact and impaired shifting

	Intact Shifting (n=58)	Impaired Shifting (n=35)	χ^2	p	Odd's ratio	Cohen's d
Anxiety Disorders						
OCD	48.3%	48.6%	0.00	0.98	1.01	0.00
OCPD	34.5%	34.3%	0.00	0.99	0.99	0.00
Panic Disorder	27.9%	36.1%	0.72	0.40	1.46	0.18
Social Phobia	36.1%	47.2%	1.17	0.28	1.58	0.23
Specific Phobia	19.7%	27.8%	0.85	0.36	1.57	0.19
PTSD	8.2%	16.7%	1.62	0.20	2.24	0.27
GAD	19.7%	13.9%	0.52	0.47	0.66	0.15
BDD	1.6%	0%	0.60	0.44	-	0.16
Multiple Diagnoses [1]	31.1%	58.3%	6.91	<0.001	2.96	0.57**
Mood Disorders						
MDD	70.5%	77.8%	0.61	0.43	1.47	0.16
Bipolar Disorder	6.6%	2.8%	0.66	0.42	0.41	0.17
Dysthymia	3.3%	2.8%	0.02	0.89	0.84	0.03
Self-harm	35.0%	58.8%	5.01	0.03 [£]	2.56	0.48*
Substance Disorders						

Alcohol Abuse	11.5%	19.4%	1.16	0.28	1.86	0.23
Alcohol Dependence	14.8%	13.9%	0.01	0.91	0.93	0.02
Substance Abuse	4.9%	13.9%	2.41	0.12	3.12	0.33
Substance Dependence	11.5%	16.7%	0.53	0.47	1.54	0.07

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder

[1] Split into those with no or one anxiety diagnosis and those with 2 or more (multiple) anxiety diagnoses

* Comparison significant at 0.05 level

** Comparison significant at 0.001 level

[£] Comparison no longer significant after Hochberg correction

Figure 17: Study 1.4 Frequency of anxiety diagnoses for women with intact shifting
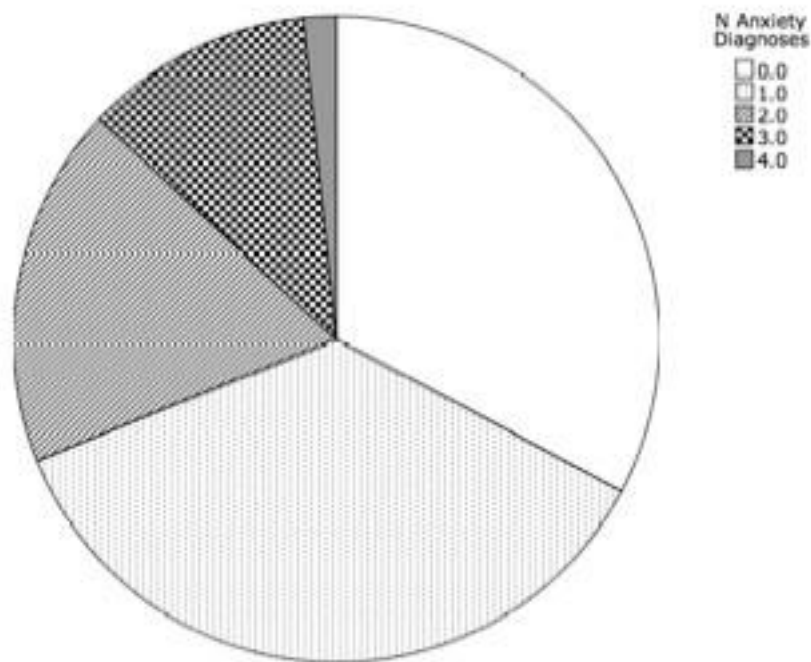
Figure 18: Study 1.4 Frequency of anxiety diagnoses for women with impaired shifting
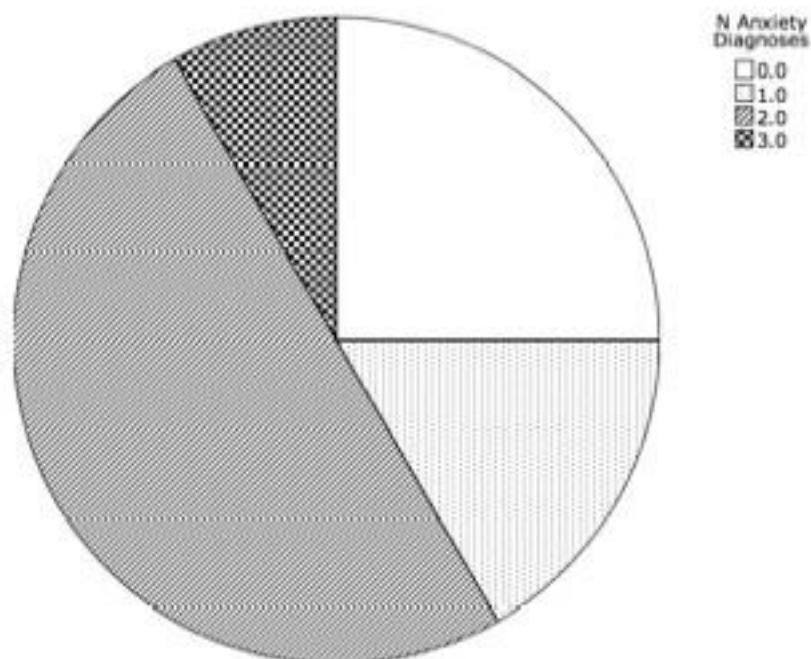
an odd's ratio just over 2.5. No significant differences were found between rates of substance disorders however those with impaired shifting were 3 times more likely to have a history of drug abuse.

6.2.8.2   Coherence results

Frequency data: Figure 19 illustrates the percentage of participants by diagnostic group with extreme scores on each neuropsychological task. The ROCF showed the highest number of extreme score cases across diagnostic groups, with BN women showing the most pronounced figures (56.7%). Pearson's chi-square test found no significant overall group difference for the GEFT ($\chi^2(3)$=3.31, p=0.35) however a difference was present on the ROCF ($\chi^2(3)$=20.81, p<0.001). Significantly more cases of extreme scores on the ROCF were found in ANR (p=0.001), ANBP (p<0.01) and BN (p<0.001) groups compared to HC.

The GEFT and ROCF did not correlate with each other, Spearman's r(91)=-0.05, p=0.63. This confirms that these two variables are not measuring a unitary concept, therefore justifying an alternative method to that employed for set-shifting to calculate the coherence composite score. As outlined in 6.2.7.2, data for both the GEFT and ROCF were split by HC quartiles. Those with AN showed the highest proportion of cases in the lower 25th quartile on the GEFT (see Figure 20). Nearly half of all ED cases fell in the lower 25th quartile on the ROCF (see Figure 21). When collapsed across task and strategy, nearly half (42.9%) of all ED cases showed persistent detail focus (ANR 46.9%; ANBP 43.8%; BN 48.1%). It is interesting to note that despite the original quartiles being created based on HC performance (therefore 25% of HC cases falling into each quartile), once collapsed a higher proportion of HC cases emerge as having good performance across both tasks (adaptive detail focus; 30.8%).

Demographic and clinical features: Current ED cases adopting a local processing style on one or both tasks (adaptive or persistent detail focus) were selected for further analysis. No significant differences emerged across demographic and clinical features (see Table 31). Effect size analysis revealed that those with persistent detail focus were moderately older, had a moderately more severe illness (at the time of testing) and a moderately later age of ED onset.

Self-report clinical features: Those with persistent detail focus were significantly more depressed with a moderate/large effect size (see Table 32,

Figure 19: Study 1.4 Percentage of participants by diagnostic group with extreme scores for each coherence task
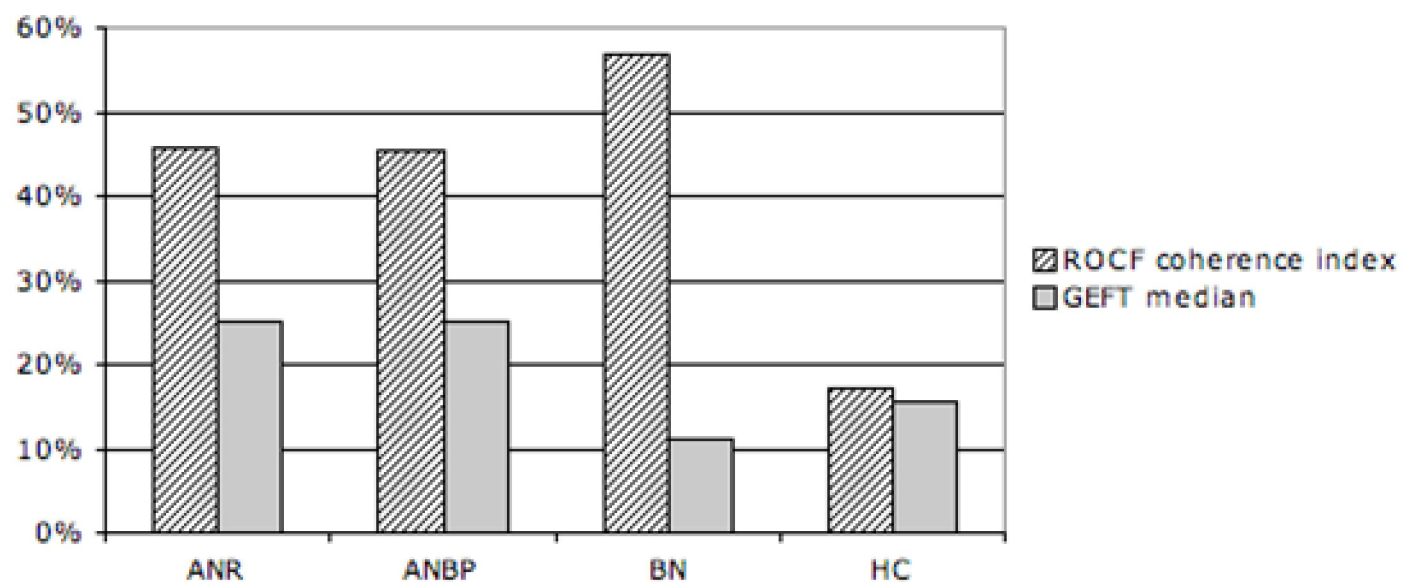
ROCF Rey-Osterrieth Complex Figure; GEFT Group Embedded Figure Test; ANR Restricting type Anorexia Nervosa; ANBP Binge/purging type Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control

Figure 20: Study 1.4 Proportion of participants by diagnostic group per quartile on the GEFT
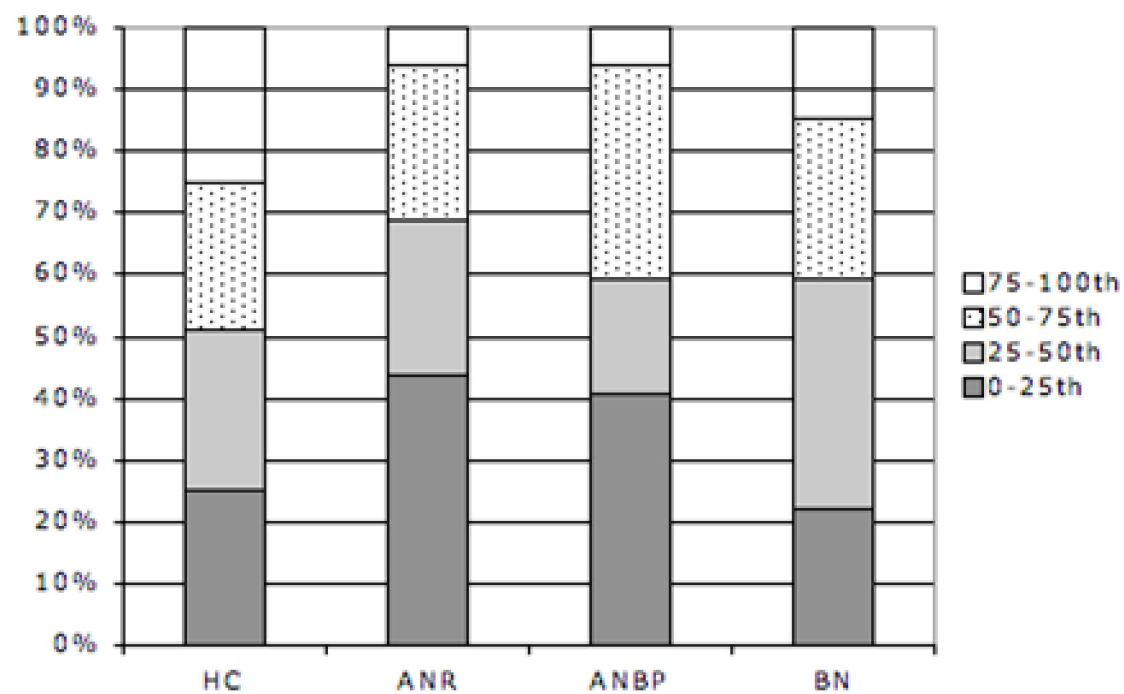
ANR Restricting type Anorexia Nervosa; ANBP Binge/purging type Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control

Figure 21: Study 1.4 Proportion of participants by diagnostic group per quartile on the ROCF
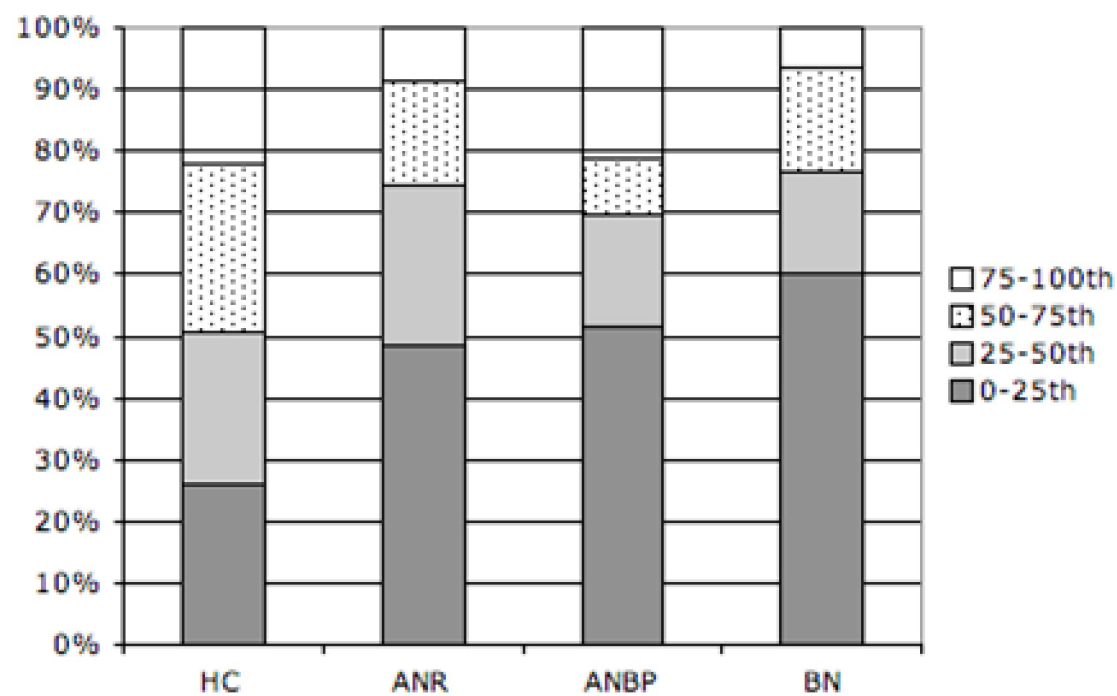
ANR Restricting type Anorexia Nervosa; ANBP Binge/purging type Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control

Figure 22: Study 1.4 Coherence strategy across both tasks by diagnostic group
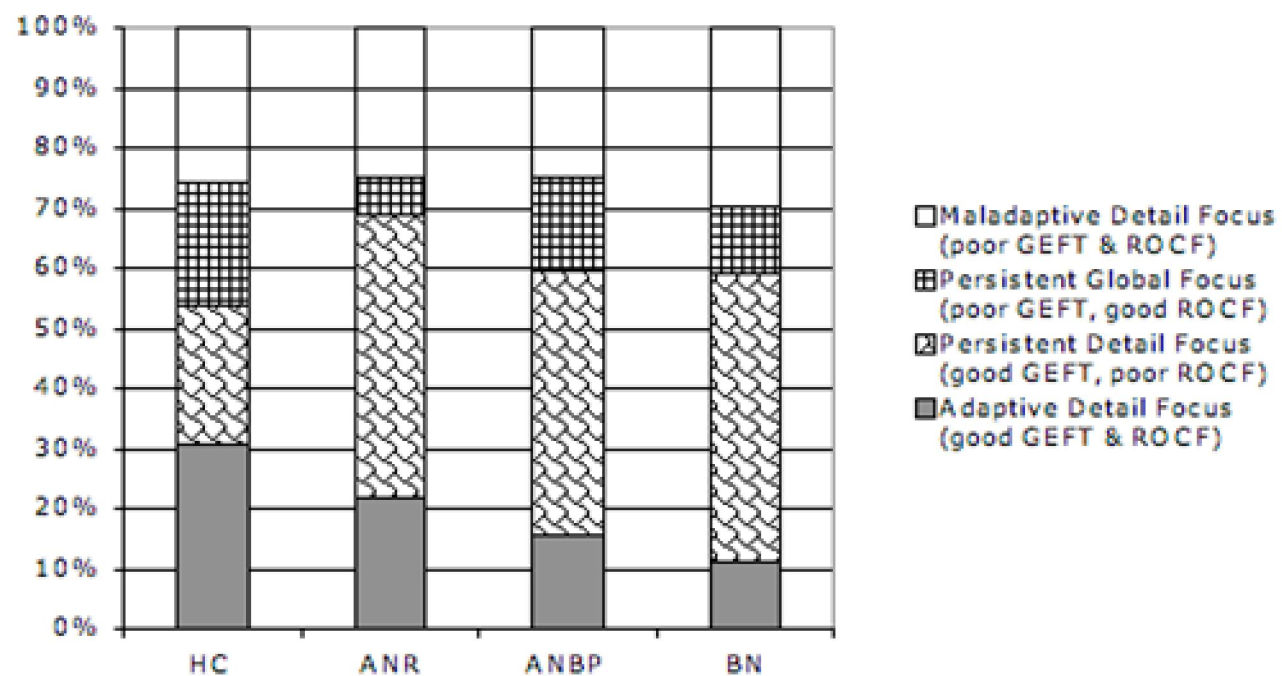
ANR Restricting type Anorexia Nervosa; ANBP Binge/purging type Anorexia Nervosa; BN Bulimia Nervosa; HC Healthy Control

Table 31: Study 1.4 Descriptive statistics for demographic and clinical features for current ED by local processing strategy

	Adaptive Detail Focus (n=15)	Persistent Detail Focus (n=42)	Test statistics			
			t	χ^2	p	Cohen's d [2]
Age	21.47 (3.31)	23.17 (4.57)	-1.32	-	0.19	0.40
Education (years)	15.53 (2.70)	15.52 (2.93)	0.01	-	0.99	0.00
Current medication [1]	0 (0-1)	0 (0-1)	-	93.0	0.61	0.18
Current ED severity [eq]	3.00 (0.66)	2.52 (1.17)	1.92	-	0.06	-0.45
Age of onset	15.00 (2.30)	16.69 (3.36)	-1.80	-	0.08	0.54
Duration of illness	6.97 (4.17)	5.76 (3.70)	1.05	-	0.30	-0.32
YBC preoccupations	12.29 (2.27)	12.63 (2.61)	-0.45	-	0.66	0.14
YBC rituals	10.50 (3.41)	11.39 (3.78)	-0.78	-	0.44	0.24
	(n=12)	(n=29)	MW			
Current BMI	18.73 (2.16)	17.96 (2.37)	0.97	-	0.34	0.33
Lowest BMI [1]	14.10 (13.20-15.60)	14.65 (13.27-16.93)	-	134.0	0.26	0.36

ED Eating Disorder; YBC Yale-Brown-Cornell Eating Disorder Scale; BMI Body Mass Index

[eq] Equal variances not assumed

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons

Table 32: Study 1.4 Self-report clinical features for current ED by local processing strategy

	Adaptive Detail Focus	Persistent Detail Focus	Test statistics			
	(n=14)	(n=41)	t	MW	p	Cohen's d
HADS anxiety	9.86 (4.20)	12.43 (4.76)	-1.79	-	0.08	0.56
HADS depression	4.43 (3.39)	7.02 (4.20)	-2.09	-	<0.05 [£]	0.65*
OCI-R total [eq]	15.71 (10.02)	20.82 (15.81)	-1.40	-	0.17	0.35
Hording	3.29 (2.97)	3.61 (3.22)	-0.33	-	0.74	0.10
Checking	2.00 (2.00)	2.44 (2.92)	-0.52	-	0.60	0.16
Ordering	3.07 (2.70)	4.09 (3.68)	-0.94	-	0.35	0.11
Neutralising	1.21 (1.48)	2.34 (2.05)	-1.33	-	0.19	0.59
Washing	1.71 (2.43)	2.29 (3.31)	-0.60	-	0.55	0.19
Obsessing	4.43 (3.69)	6.05 (4.32)	-1.25	-	0.22	0.39
Rosenberg self-esteem	14.21 (5.92)	12.30 (5.07)	1.17	-	0.25	-0.36
Frost Perfectionism	98.09 (18.98)	97.27 (14.89)	0.17	-	0.87	-0.05
Concern mistakes	31.36 (10.16)	32.70 (8.30)	-0.49	-	0.63	0.15
Personal standards	28.73 (4.69)	27.45 (4.90)	0.85	-	0.40	-0.26
Doubting	13.71 (4.34)	13.96 (3.22)	-0.23	-	0.82	0.07
Organisation	24.29 (6.83)	23.16 (5.18)	0.65	-	0.52	-0.20
CHiRP total [1]	8 (5-11)	7 (4-11)	-	163.0	0.61	0.15

Perfectionism [1]	3 (2.25-4)	3 (1-4)	-	161.5	0.23	0.36
Inflexibility [1]	2 (1-3)	2 (2-4)	-	167.0	0.15	0.43
Order/symmetry [1]	1 (0-3.25)	1 (0-4)	-	259.0	0.88	0.06

HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire;

[eq] Equal variances not assumed (Levene's test for equality of variance <0.05)

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

* Comparison significant at 0.05 level

[£] Comparison no longer significant after Hochberg correction

Figure 23: Study 1.4 Current depression levels in current ED by local processing style
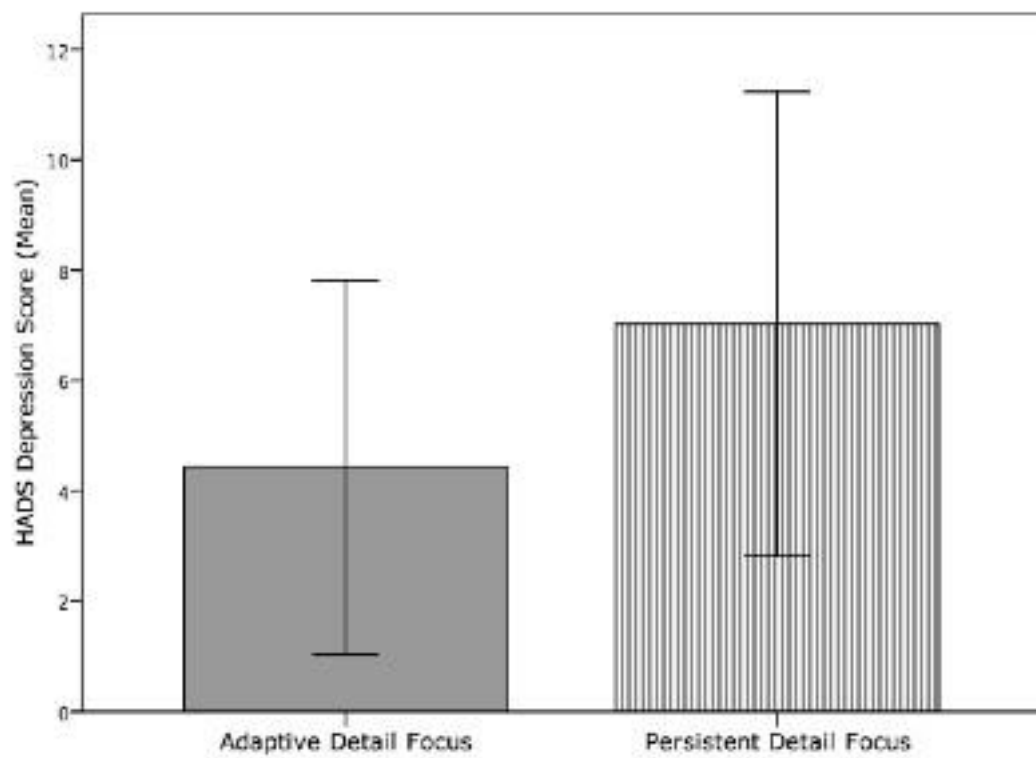
Figure 23). Women with persistent detail focus also showed a trend toward moderately increased anxiety levels, obsessive-compulsive behaviours (particularly neutralising and obsessing), and childhood inflexibility. Small/moderately lower levels of self-esteem and higher levels of childhood perfectionism were also seen.

Comorbidity: Those with persistent detail focus had significantly higher rates of social and specific phobia (see Table 33). The odds ratio for specific phobia was highest across all comorbid conditions, with those with persistent detail focus over 5 times more likely to have specific phobia. Women with persistent detail focus were nearly four times as likely to have multiple anxiety diagnoses than women with adaptive detail focus. Nearly half (46.7%) of those with adaptive detail focus had no comorbid anxiety disorder, compared to only 22% of those with persistent detail focus. Figure 24 and Figure 25 illustrate the number of anxiety diagnoses or participants across local processing groups.

Frequency of depression diagnosis was again high across both groups, with a negligible difference between them. Effect size analysis revealed a moderate effect for Dysthymia where those with persistent detail focus were less likely to endorse this diagnosis (0 cases). A moderate effect was also seen for alcohol dependence, where women with persistent detail focus were more likely to have alcohol dependency (0 cases in the adaptive detail focus group).

6.3   Discussion

The aim of this chapter was to investigate set-shifting and coherence performance of women with ED from three diagnostic perspectives; a transdiagnostic approach (i.e. all ED compared to HC women), traditional DSM-IV diagnostic split (i.e. AN and BN compared to HC), and a lifetime phenotype split (i.e. ANR, mixed ANBN, and pure BN compared to HC). Overall, support was provided for the hypothesis of set-shifting meeting criterion 1 of an endophenotype in that women with a diagnosis of ED in general displayed a more rigid or inflexible profile compared to HC women. This profile seemed more pronounced across tasks for women with a mixture of both AN and BN behaviours. The hypothesis for weak coherence was confirmed in the AN sample across tasks, in that women with AN displayed a more detail focussed processing style than HC women. Mixed findings were found for the BN group, who displayed difficulties with global processing but normal levels of detail focus. When cases were split by normative (HC) performance

Table 33: Study 1.4 Diagnostic comparisons between current ED by local processing style

	Adaptive Detail Focus	Persistent Detail Focus	χ^2	p	Odd's ratio	Cohen's d
	(n=14)	(n=40)				
Anxiety Disorders						
OCD	42.9%	60.0%	1.23	0.27	1.40	0.31
OCPD	30.8%	25.0%	0.17	0.68	0.80	0.11
Panic Disorder	33.3%	29.3%	0.09	0.77	0.88	0.08
Social Phobia	13.3%	46.3%	5.12	0.02 [£]	3.48	0.65*
Specific Phobia	6.7%	34.1%	4.23	0.04 [£]	5.12	0.58*
PTSD	6.7%	4.9%	0.07	0.79	0.73	0.07
GAD	6.7%	22.0%	1.75	0.19	3.29	0.37
BDD	0%	0%	-			
Multiple Diagnoses [1]	13.3%	51.2%	6.51	0.01 [£]	3.84	0.74*
Mood Disorders						
MDD	73.3%	78.0%	0.14	0.71	1.06	0.10
Bipolar Disorder	13.3%	4.9%	1.18	0.28	0.37	0.30
Dysthymia	6.7%	0%	2.78	0.10	-	0.47
Self-harm	40.0%	52.5%	0.68	0.41	1.31	0.23
Substance Disorders						

Alcohol Abuse	6.7%	17.1%	0.97	0.32	2.56	0.27
Alcohol Dependence	0%	17.1%	2.93	0.09	-	0.48
Substance Abuse	13.3%	7.3%	0.49	0.48	0.55	0.19
Substance Dependence	6.7%	12.2%	0.35	0.55	1.83	0.16

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder

[1] Split into those with no or one anxiety diagnosis and those with 2 or more (multiple) anxiety diagnoses

* Comparison significant at 0.05 level

£ Comparison no longer significant after Hochberg correction

Figure 24: Study 1.4 Frequency of anxiety diagnoses for women with adaptive detail focus
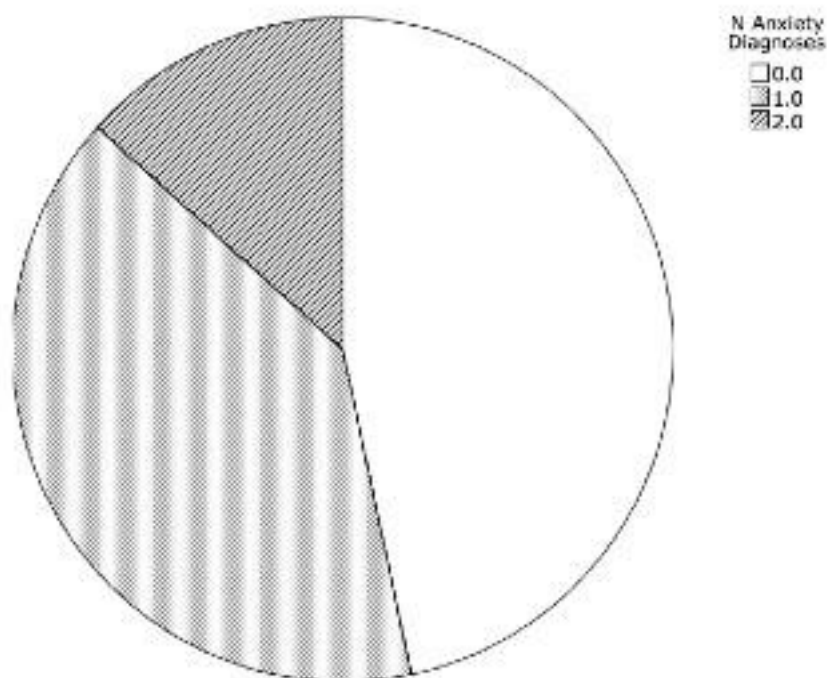
Figure 25: Study 1.4 Frequency of anxiety diagnoses for women with persistent detail focus
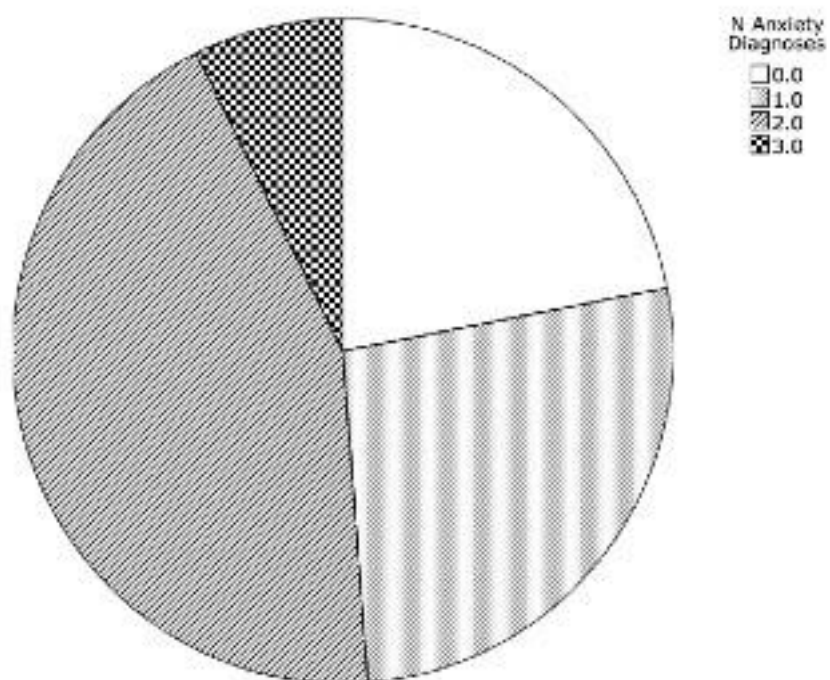
on set-shifting and coherence tasks, some interesting differences across demographic, clinical features and comorbid psychiatric diagnoses were seen.

6.3.1 Set-shifting

6.3.1.1 Analysis 1: Transdiagnostic split

Significant differences between women with and without an ED were found for at least one outcome across four of the five neuropsychological tasks employed: TMT, WCST, Brixton and Haptic tasks. Women with an ED were consistently poorer at set-shifting on these tasks, indicating a more inflexible cognitive style.

A significant difference on the TMT was found for the raw shift time, but not for the balanced (B-A) time variable or the error count variable. A small effect was seen across all three outcomes. The conclusion of impaired set-shifting based on findings from the TMT must therefore be made tentatively. Despite the literature most often reporting raw time as a measure of set-shifting (Roberts et al., 2007b), particularly when employing a computer version of the TMT the raw shift time is substantially influenced by extraneous factors such as a participants capability moving a mouse on the computer screen. This means that a flexible individual with poor (slow) mouse skills may get the same raw TMT time as an inflexible individual with excellent (fast) mouse skills. Therefore it is proposed that the balanced B-A variable reported here provides a more refined measure of set-shifting ability, as to a degree it is able to take baseline mouse speed into account. The difference between groups on the balanced variable falls just short of significance, however descriptive statistics trend in the same direction as that of the raw time variable where women with ED show a larger difference in time between the shift trial (B) and the control trial (A). This indicates that they may be slower therefore poorer at shifting between letter and number sorting (when accounting for baseline time) than are HC women.

Significant differences between women with and without an ED were seen on both outcomes of the WCST (perseverative errors and number of categories completed) and for both the Brixton and Haptic tasks. Across all variables, the ED group were poor on set-shifting, as evidenced by more perseverative errors and fewer categories completed on the WCST, more errors made on the Brixton task, and more illusions experienced on the Haptic task. This reflects a more inflexible cognitive style in the ED group. The size of the difference (based on effect sizes) was small for Brixton and Haptic tasks, but moderate and large for the WCST, suggesting that the

WCST is the most sensitive outcome measure for set-shifting impairment within the current task battery.

No differences between groups were found for any of the three outcome variables for the CatBat task; raw time, balanced (Bat minus Cat) time, or number of errors made. Negligible to small differences between group are seen across variables. This could indicate either a) a lack of sensitivity of the CatBat task in the present sample, or b) any effects could be masked due to the heterogeneity of the current ED group. Power analysis determined that the CatBat task required the largest sample size (n=40) of all neuropsychological tasks to detect between group differences. Both groups employed a sample size well over this number, therefore the lack of a significant result is unlikely due to a lack of statistical power.

The largest effect sizes in this study were seen for self-report cognitive flexibility as measured by the CFS and TSQ. The differences between clinical and control groups on these measures are substantially larger than that observed on neuropsychological measures, indicating one of two possibilities: either self-report is a more sensitive measure of cognitive flexibility, or self-report and neuropsychological assessment measure different concepts. The nature of self-report is subjective, in that an individual requires a degree of insight into his or her own thoughts and behaviours in order to rate each item. Given that questions regarding thinking style ask an individual to rate aspects of their behaviour and cognition (e.g. "I avoid new and unusual situations"), these ratings could be easily confounded by factors such as depression, anxiety and self-esteem. Indeed, subsequent analysis found both the CFS and TSQ to be on average strongly correlated with self-report depression ($r(91)=-0.60$, $p<0.001$; $r(94)=0.48$, $p<0.001$), anxiety ($r(91)=-0.57$, $p<0.001$; $r(94)=0.51$, $p<0.001$), and self-esteem ($r(92)=0.67$, $p<0.001$; $r(95)=-0.35$, $p<0.01$) in women with current ED. By contrast, neuropsychological assessment is an objective measure of specifically targeted concepts, where performance is unlikely to be dramatically influenced by these extraneous factors. In order to empirically test the second possibility, self-report flexibility and neuropsychological assessment were correlated. Only the TMT correlated significantly with both the CFS and TSQ, indicating some overlap but not to the extent that would be expected if they were tapping the same concept. These findings suggest that subjective and objective cognitive flexibility are not particularly well matched, where the former is greatly influenced by an individual's state and self-perception.

All current ED diagnoses were assessed as one group in this analysis. Heterogeneity within this clinical group may help explain the wide variance across some neuropsychological results, causing subtler group differences to be lost. This hypothesis will be discussed in the following sections.

6.3.1.2   *Analysis 2: Traditional diagnostic split*

As outlined in the results section for analysis 2 (see 6.2.4), significant differences were found between AN and/or BN and HC groups across the TMT, WCST, Brixton and Haptic tasks. Across all four tasks, women with current ED were less flexible than the HC group. Results for the AN group indicated poorer shifting on WCST and Brixton tasks, with a trend toward more illusions on the Haptic task. These findings support the set-shifting hypothesis for the AN group, indicating difficulties in cognitive flexibility compared to HC. No differences between AN and HC groups were found for the TMT and CatBat tasks.

For the BN group, significant differences were found on the TMT, WCST and Haptic tasks compared to HC. No difference was found for the Brixton task, where a negligible effect size was seen, or on the CatBat task as with the AN group. These results also provide support for the hypothesis of set-shifting difficulties in the BN population compared to HC.

Overall, the WCST was the only task to show robust group differences across clinical and control groups. This was especially evident for the AN group, who made nearly twice the number of (mean) perseverative errors as HC and significantly more errors than both HC and BN groups. The BN group showed a trend toward more perseverative errors however scored significantly different to HC on the WCST only on the categories completed variable. Raw perseverative errors is a more pure measure of set-shifting ability, as failure to complete all six categories could be as much influenced by random errors as by difficulties shifting set. Therefore results from the WCST indicate heightened difficulties with set-shifting in the AN compared to BN group. This is further supported by results from the Brixton task, where the AN group only show impaired flexibility (more errors) compared to HC.

In contrast, results from the TMT are suggestive of the opposite pattern, with BN taking significantly longer to complete the shifting component of this task compared to HC, in addition to making more errors. While it is readily acknowledged that no task of executive functioning taps into a single construct such as set-shifting (hence the use of a battery of tasks), one advantage of the WCST is

that multiple outcome variables detail a number of factors such as perseverative errors versus random errors, in addition to variables such conceptual level responses and a score of rule learning ability. This allows the investigator to choose the outcome most pertinent to their research question, in this case perseverative errors. The TMT does not allow for such a sophisticated breakdown of results, and subsequently one can be less certain about the degree to which extraneous factors play a role in task outcome. Three previous studies to date have used the TMT in a BN population, and while two report findings consistent to those presented here (Pendleton-Jones et al., 1991; Tchanturia et al., 2004a), one reports the opposite (Murphy et al., 2002). Study sample sizes are not large at 38, 18 and 16 (respectively) but not dissimilar to the current study's 30 BN participants. It may be that BN specific tendencies such as heightened impulsivity (illustrated by more task errors) and inattention (Rosval et al., 2006) delayed their completion of the task rather than poor set-shifting. Indeed, the TMT is not purely a measure of flexibility, but is also used to measure working memory and aspects of attention (Lezak et al., 2004). Using the balanced TMT outcome (B-A) goes some way toward controlling for the attention aspect of the task, however it cannot be assumed that this variable would discount attentional demands altogether. The role of attention on TMT performance cannot be readily tested in the current sample as no independent impulsivity or attentional measure was included.

It may also be that the AN group simply found the TMT particularly easy compared to the WCST and Brixton task. This could be because of either less difficulty with the shifting aspect of the task, or less difficulty with the attentional aspect of the task. The TMT was consistently presented as the first task in the battery, and it is possible that those with AN were more able to focus their attention on task completion at this initial stage than those with BN were. The TMT has been more widely used in the AN compared to BN population, where all but one investigation by Tchanturia et al. (2004a) report no significant difference between AN and HC groups (Pendleton-Jones et al., 1991; Kingston et al., 1996; Mathias & Kent, 1998; Murphy et al., 2002; Holliday et al., 2005; Steinglass et al., 2006), contributing small effect sizes to the TMT meta-analysis (Roberts et al., 2007b). These studies employed a large range of clinical participants from severely underweight (BMI<15) to largely weight restored as in the current study (BMI 17-18). Given the severe nature of the AN inpatient sample employed by Tchanturia et

al., it may be that in such a starvation state participants struggled with the attention aspect of the task rather than its shifting component. It is suggested that the TMT may be a task more sensitive in the BN rather than AN population and perhaps more sensitive to attentional demands than was previously acknowledged in the ED literature. Further research to include parallel tasks of impulsivity and attention are required in order to further understanding of the mechanisms of the TMT.

The Haptic task provides evidence for impaired set-shifting in both AN and BN, with small effect sizes for both groups compared to HC, reaching significance for BN only. AN and BN groups did not differ dramatically compared to each other, with only a small effect size (d=0.17). The Haptic task again is not a clean measure of set-shifting and perhaps is most susceptible to extraneous factors of all tasks administered in the current study given its reliance on sensory processing. It has been suggested in the previous meta-analysis chapter (see 4.4) that large effect sizes on the Haptic task indicative of substantial set-shifting impairment may be inflated due to basic deficits in tactile processing in the AN group as found by Grunwald and colleagues (Grunwald et al., 2001a; Grunwald et al., 2001b). Only one study could be found investigating tactile processing in the BN population where no deficit was found (Faris, Raymond, De Zwaan, Howard, Eckert, & Mitchell, 1992), making the current finding of a larger effect in the BN group difficult to interpret. Perhaps most notable regarding the descriptive results for the Haptic task is the number of illusions experienced by the HC group, scoring nearly twice that of the HC samples previously reported (Tchanturia et al., 2002; Tchanturia et al., 2004a; Holliday et al., 2005). This is likely why the Haptic task, usually a robust measure with a large effect size (Roberts et al., 2007b) fell to small effects in the current study. Indeed were the current AN group compared to the HC group of Tchanturia et al. (2004a), the effect size of 0.30 would increase to a very large 1.16. This illustrates a higher level of perceptual rigidity in the current HC sample than has previously been found.

No differences were found on the CatBat task between HC and either AN or BN groups. Inspection of descriptive statistics across studies employing the CatBat task highlights comparable results for clinical groups, but like the Haptic task a large discrepancy in HC performance. For example, in comparison with the Tchanturia et al. (2004a) sample, HC performance is nearly twice as slow thus masking the consistent result of the clinical group. Again, were the current AN sample compared to the HC group reported in the Tchanturia paper, the effect size would change

markedly from negligible (0.11) to large (1.03). The inflation of the present HC sample's Bat time is also evident in comparison to other studies reporting a HC sample for this task (Tchanturia et al., 2002; Holliday et al., 2005). This illustrates that the lack of a significant finding for the CatBat task in the present thesis may be more a factor of impaired performance or task effects in the current HC group than a lack of difficulty with this task in the ED groups.

6.3.1.3 Analysis 3: Lifetime Phenotypic Split

Analysis 3 split the clinical group by lifetime phenotype, where ANR, mixed ANBN and pure BN groups were compared with HC women. As in the previous analyses, significant differences were found between clinical and control groups on the TMT, WCST, Brixton and Haptic task, with the current diagnostic split also showing one significant post-hoc comparison on the CatBat task. Significant differences were distributed across clinical groups. ANR showed significantly poorer flexibility on the WCST and Haptic tasks, with no difference compared to HC on the Brixton (as was reported with the combined AN group in analysis 2). The mixed ANBN group showed poor flexibility on the TMT (raw time only), WCST and Brixton tasks, while the BN group differed on the TMT, WCST (categories completed only) and CatBat errors. As previously reported, the WCST was the only task to show consistent significant differences across diagnostic groups, despite the small sample size of the BN group (n=13).

On top of the traditional diagnostic split reported in analysis 2, the current split by phenotype further informs group differences on a number of tasks. The sensitivity of the TMT in the BN but not AN population persisted with the current phenotypic split, where the most robust differences were found in the pure BN group. While the ANBN group initially showed longer raw latencies on trail B, this difference disappeared when baseline mouse speed (a measure of attention) was accounted for (B-A variable). This shows that attention is indeed a factor in TMT performance, as on controlling for attention (to some degree) with trail A time, the purported set-shifting impairment in the ANBN group on this task disappeared. This finding is consistent with previous experimental work on attention and impulsivity in ED, where a deficit was found in BN and ANB/P (i.e binging ED) but not ANR on the Go-NoGo task and several self-report measures of attention and impulsivity (Rosval et al., 2006). These results suggest intact attention (and therefore no difference in raw trail B time) in the ANR group, but poor attention (and therefore

slower trail B time) across binging groups. This further supports the previously suggested investigation of TMT performance in parallel to measures of attention and impulsivity.

Secondly, robust findings on the WCST in AN spectrum disorders and to a lesser extent in BN spectrum disorders is replicated with the current split. Small numbers of pure BN cases must be noted, which may explain the non-significant comparison in spite of a large effect size between BN and HC. Investigation of descriptive statistics suggests that the pure BN group performed better at shifting on the WCST across both outcome measures compared to AN groups. Contrary to expectation, the finding of significantly more errors on the Brixton task in the AN group persisted for the ANBN mixed cohort, but not for ANR despite the two groups differing by an average of only 1.25 errors. It should be noted that across both AN groups, raw number of errors were substantially less (11/12 compared to 17) than has been previously found on this task in current AN (Tchanturia et al., 2004a). Even so, a lack of significant group differences on the Brixton has been reported before in the ED population (Holliday et al., 2005). It is possible that task effects could have played a role in the lessened number of errors reported here: the Brixton task was administered third in the neuropsychological battery, following the TMT and WCST. It is possible that by this stage, the clinical participant had been primed to the practise of cognitive flexibility, and thus performance more approximated 'normal' compared to if the Brixton task had for example been presented first. In order to avoid this possibility in future studies, careful thought should be given to the order of tasks, preferably so that tasks assessing the same construct do not directly follow each other.

Thirdly, the slightly puzzling finding of a more illusions (errors) on the Haptic task in the BN group in analysis 2 was not replicated with the current phenotypic split, whereas the trend toward more illusions for those with AN became more prominent with the separation of ANR and ANBN. The ANR group showed significantly more illusions than HC on the Haptic task with a moderate effect size. As discussed above, difficulties with tactile processing per se in the AN group (Grunwald et al., 2001b) may exaggerate the results seen here in the context of poor cognitive flexibility. Despite this, results suggest more prominent perceptual shifting deficits amongst pure restricting AN.

Finally, though no notable results on the CatBat task were found in the previous two analyses, a significant finding on number of errors emerged with the current group split. A significant post-hoc test was found between BN and HC groups, where the BN group made more errors in the Bat half of the story than did HC, indicating more perseverations and therefore poorer set-shifting. Given the lack of a difference in time variables for the CatBat task, it is possible that the higher number of errors on the CatBat task in the BN group could again be partly due to impulsivity rather than set-shifting. As previously outlined, this notion requires further investigation with parallel measures of attention and impulsivity before conclusions can be made. Additionally, given the large sample size disparity in the pure BN (n=17) compared to HC (n=88) group, these results should be interpreted with caution.

6.3.1.4 *Analysis 4: Extreme scores*

In analysis 4 both clinical and control cases were separated into those with extreme scores (± 1 SD from HC mean) and not, across the four set-shifting tasks (TMT, WCST, Brixton task & Haptic Illusion). Data from these tasks was then integrated by creating a composite variable, where those with extreme scores on two or more tasks were categorised as having impaired shifting, and those with extreme scores on one or no tasks were categorised as having intact shifting. Only a small number of HC women displayed poor shifting (just over 10%), suggesting that notable difficulties with cognitive flexibility are relatively uncommon in the general population. By comparison, impaired shifting is seen in between 25-50% of current ED cases, with the highest rate found in the ANBP group. All current ED groups (ANR, ANBP, BN) had a significantly higher number of impaired shifting cases compared to HC. As this is the first study of set-shifting to investigate AN subtypes, it is interesting to note the trend toward higher levels of impaired shifting in ANBP compared to ANR cases.

A number of significant findings emerged when comparing demographic and clinical variables across all current ED cases split by impaired or intact shifting. Those with impaired shifting had a longer illness duration than those with intact shifting. Additionally, a more severe illness at its worst in terms of ED related rituals as measured by the YBC-EDS was seen in the impaired shifting group. For example, those with impaired shifting reported higher distress and more time spent carrying out food rituals such as computing calorie/fat content of food and cutting food into a

specific size, body rituals such as ritualised weighing, spanning their wrist with fingers or checking that their hip/collarbone is visible, and/or exercise rituals such as the need to exercise in a specific way/a specific amount or needing to shiver instead of putting on additional clothing. Evidence that poor set-shifting is independent of illness state and BMI is provided, as negligible differences were seen on these variables across set-shifting ability. These findings suggest that those with poor cognitive flexibility are more likely to respond poorly to treatment, as their difficulty changing cognitive processes or behavioural patterns causes eating related behaviours to become reinforced and more pervasive over time. This difficulty initiating or implementing change results in a longer illness. The absence of a difference in illness state or current BMI across intact and impaired shifting groups provides further support for this trait as a state-independent endophenotype, rather than a factor of starvation.

In addition to illness related variables, significantly higher numbers of comorbid anxiety disorders were seen, along with higher rates of self-harm and lower levels of self-esteem. Comorbid anxiety disorders are known to be associated with poor outcome in AN (Steinhausen, 2002)., and low self-esteem has been found to predict poor treatment adherence (Halmi et al., 2005). All of these factors point toward poor prognostic indicators being associated with impaired shifting. These findings have pertinent treatment applications, which will be further discussed in the final chapter.

Previous research on cognitive flexibility in ED has often reported illness features and levels of comorbid symptomology (usually through self-report measures), however such variables have been largely presented as clinical descriptors of the sample. Some papers have included for example self-report depression and obsessionality or BMI as covariates of neuropsychological performance (Tchanturia et al., 2004a; Wilsdon & Wade, 2006) however no association between these variables and set-shifting performance has been found. Similarly, no significant correlations between set-shifting scores and self-report depression, anxiety, and BMI have been found (Mathias & Kent, 1998; Tchanturia et al., 2004a; Fowler et al., 2005). A small but significant correlation has been reported between age and a composite variable consisting of the TMT and Brixton tasks (Tchanturia et al., 2004a), however no relationship with age was found in the current study. In this thesis, a dichotomous rather than correlational approach was employed.

This, in addition to the large sample size in both clinical and control groups (up to four times that of previous studies) may have allowed for greater power to detect differences in measures of comorbidity and cognitive performance.

The composite variable has allowed for a more simplified analysis of the data by allowing results to be collapsed across tasks and their various normative scores. A clear picture of impaired cognitive flexibility across ED groups emerged.

6.3.2   Coherence Discussion

6.3.2.1   Analysis 1: Transdiagnostic split

Significant differences were found between ED and control groups on the two main coherence outcome measures: ROCF copy central coherence index, and GEFT median time. A large effect size was seen for ROCF across both the index total and order and style indices, where women with ED prioritised detail and displayed a more fragmented drawing style then HC women. Likewise, on the GEFT women with an ED were significantly faster, and made fewer time out errors than women without an ED, indicating that they were less likely than HC to fail in finding a shape within the 60 second time-frame. Results from both tasks support the weak coherence hypothesis, in that women with ED display a more detail focussed processing style compare to HC women, who are more global in their approach. This replicates findings from previous studies, who have also found a superiority with detail focussed processing in those with ED (Tokley & Kemps, 2007; Lopez et al., 2008c). It is interesting to note the difference in effect size between the two tasks. While a large effect size was seen for ROCF coherence index, a small effect size is found for GEFT. Investigation of the data spread for each task shows a very tight range for the HC group on the ROCF, with a substantially lower 25th quartile for the ED group. The range for the ED group is more than twice as wide as the HC group. In comparison, for the GEFT a wider range is seen in the HC group rather than the clinical group, with a range nearly twice that of the ED group. Previous studies using the GEFT have employed slightly different methodology, reporting means and standard deviations that showed a more similar range across clinical and control groups and a large effect size (Tokley & Kemps, 2007). The heterogeneity in the current HC group for GEFT median time may help explain the decreased sensitivity of the GEFT task and thus small effect size.

Copy and recall accuracy for the ROCF showed the expected pattern of results based on previous research. No difference was found between clinical and control groups for copy accuracy. This is consistent with previous findings for AN, where despite employing a different organisational strategy (as measured by the central coherence index) no difference in the accuracy of the drawing was found (Lopez et al., 2008b). Those with BN have been found to show significantly lower accuracy compared to HC (Lopez et al., 2008d) however as the current ED sample presented here is over two thirds AN it is not surprising that results are in line with previous AN findings. Also consistent with the previously mentioned studies, the current ED group scored significantly lower on recall accuracy compared to HC, and a positive correlation between copy central coherence index and recall accuracy was found. This correlation indicates that those employing a more global drawing style when copying the drawing are more likely to remember the figure better after 20 minutes. As prior research has investigated this relationship amongst diagnostic groups rather than using a mixed cohort (as is discussed here), copy and recall accuracy scores will be explored further below.

6.3.2.2 Analysis 2: Traditional diagnostic split

Analysis 2 split the clinical group by lifetime DSM-IV diagnosis of AN or BN. Results from both the ROCF and the GEFT indicated that those with current AN displayed a bias toward detail focussed processing compared to HC. On the ROCF, the AN group both prioritised the drawing of local elements, in addition to adopting a more fragmented style compared to HC. This is indicative of a bias toward detail in the AN group. On the GEFT, AN were both significantly faster than HC and also made significantly fewer time out errors. This further suggests a superiority in detail focused processing across tasks, confirming hypothesis 1 with regard to coherence in AN and replicating findings from previous research in this area (Southgate et al., 2007; Tokley & Kemps, 2007; Lopez et al., 2008b). Regarding the BN group, low scores were found on the ROCF like those of the AN group, suggesting a bias toward detail and/or poor global integration. A bias toward detail was not confirmed on the GEFT, where results from the BN group were similar to HC. This seemingly inconsistent finding between tasks may be suggestive of a deficit in global processing but a normal level of detail focus in those with BN.

Measurement of the organisational strategy of the ROCF is a relatively recent endeavour in the ED population, with only three studies (2 AN; 1 BN) reporting this

assessment in the literature to date (Sherman et al., 2006; Lopez et al., 2008b; Lopez et al., 2008d). Studies by Lopez et al. assessed organisation strategy using the same "coherence index" (Booth, 2006) employed in the current thesis, while Sherman et al. employed the "Savage" system, roughly synonymous to the style measure of the coherence index. Both systems assess priority of detailed over global processing. The significantly lower coherence score for AN compared to HC in the current study is consistent with both the Sherman and Lopez AN studies, as indeed is the finding of a lower coherence index in the BN group compared to HC (Lopez et al., 2008d). This illustrates a consistent finding of detail focus and/or poor global integration across both AN and BN on the ROCF with the available literature.

Previous research has failed to fully understand which aspects of the weak coherence hypothesis the ROCF is measuring. While the coherence index is an obvious measure of detail focus, the ROCF also benefits from global integration (in terms of accuracy) therefore making it a task tapping both aspects of the weak coherence hypothesis. It is suggested here that the coherence index should be interpreted as a measure of superiority with detail or as a measure of poor global integration based on the context of accuracy scores: The general population completes the ROCF using a global strategy, and scoring high on accuracy. Thus the conclusion is that they have intact global processing, and do not display a bias toward detail. The BN population completes the ROCF using a fragmented drawing strategy, but with significantly poorer scores on accuracy. Low accuracy when copying suggests poor global integration, which in this context can also explain the fragmented drawing style. Therefore for the BN population, the low coherence index is indicative of poor global integration rather the detail focus. In comparison, the AN population completes the ROCF using the same fragmented strategy as the BN group, but scores high on accuracy. This high accuracy score suggests intact global integration in the AN group. Therefore, the fragmentation of the drawing is due to an inherent bias toward detail. Consistent with the new conceptualisation of the weak coherence account (Happe & Booth, 2008), this indicates that the AN group have a natural bias toward detail that does not interfere with their ability to see the bigger picture (see Appendix 9 for example ROCF drawings from AN, BN and HC participants).

In order to formally test the hypothesis of poor global integration (i.e. low copy accuracy) contributing to the low coherence index score, the ROCF coherence

index analysis was retrospectively re-run with ROCF copy accuracy (as a measure of global integration) as a covariate. Copy accuracy was a significant covariate ($p<0.01$), indicating that that it plays a marked role in coherence score. Additionally, across all clinical groups there was a positive correlation between ROCF copy accuracy and central coherence index ($r(98)=0.33$, $p=0.001$), indicating that those with lower drawing accuracy were likely to have lower central coherence scores, although it is difficult to conclude the direction of influence. This mechanism merits further investigation, for example it may be pertinent to time how long each participant takes to copy the ROCF, in order to further investigate the impact that time/care taken in drawing construction has on the central coherence index.

Some additional factors may influence the low coherence index for BN. It is possible that aspects of the BN profile such as heightened impulsivity, lack of attention and/or planning could further play a role in the organisation of the drawing, contributing to the low central coherence index. A pilot study investigating weak coherence in the overweight/obese population indicated an even more pronounced deficit in copy accuracy than that found in BN (Roberts, Dermetriou, Treasure, & Tchanturia, 2007a). Low scores on the central coherence index are also reported in the overweight group, where qualitative reports indicate a notable lack of planning and organisation with regard to figure copying, with the overweight group described as employing a "chaotic" drawing style, with little or no regard for the task instruction to copy the figure as carefully as they can. At the opposite end of the spectrum, the current AN participants in this thesis are observed as slow and meticulous with this task, clearly following the task instruction and, in numerous cases, even asking for a ruler. It is possible that some elements of this impulsive, inattentive style characteristic of BN (also seen in the overweight population) is contributing to poor global integration, lowering accuracy scores which in turn has a direct impact on central coherence score.

Further support for the conclusion of enhanced detail focus in the AN but not BN population is provided by GEFT results, where those with AN are faster at identifying local stimuli (details) within a larger more complicated shape or context. This suggests a superiority with detail focussed processing, and compliments the ROCF findings discussed above. In contrast, a negligible effect size is found on the GEFT for the BN group, where performance is comparable to the HC group on both median time and time out errors. As the GEFT is a more pure measure of detail

focus, this further supports the conclusion from ROCF findings that those with BN do not show superiority with detail.

Despite the coherent findings for the BN group across tasks in this thesis, previous research has found similar results for BN as for AN on the Embedded Figures Test (EFT). Lopez et al. (2008d) administered the EFT to 42 women with BN and 42 HC. Women with BN were significantly faster than HC with a moderate to large effect size (d=0.69). Comparison of the descriptive statistics reveals that the current BN group had a median time only 1.2 seconds slower than the BN group of Lopez et al. (2008d), however the current BN group did not differ significantly from HC while the Lopez BN cohort did. This discrepancy could be in part due to a number of methodological differences: the GEFT employed in this thesis consisted of 18 test shapes compared to only 12 shapes in the Lopez study. It is possible that the use of additional shapes may have affected concentration levels in the BN group, contributing to the slower time in the current study. The nature of the task requires a large degree of attention, where even a few seconds of distractibility whilst searching for a shape could form the time discrepancy between AN and BN scores. Additionally, while the GEFT shapes are shaded with blue and white only, the EFT is a full colour assessment, which may make identification of target shapes harder. Such explanations are unlikely to make a large difference, and any effect should have influenced both clinical and HC groups. It is however noteworthy that the HC group in this thesis performed nearly 5 seconds (or 40%) faster than the HC group in the Lopez studies (Lopez et al., 2008b; Lopez et al., 2008d), thus removing a large portion of the previously reported time difference between the groups. This disparity across HC groups could help explain the lack of a significant finding in the current BN group. Were the current BN group compared to the HC group of Lopez et al., a moderate effect size would have been seen. Given that this is only the second study to report on weak coherence in the BN population, further replication taking into account variables such as attention and impulsivity is required before more concrete conclusions can be drawn.

One further point is of note regarding accuracy scores on the ROCF. Rather than look at group differences, the purpose of ROCF recall accuracy was to investigate whether there was a relationship between copy central coherence index and memory accuracy at 20-minute spontaneous recall. It was hypothesised that those with a more global processing style (higher central coherence index for copy)

would have used a more adaptive information processing style for the task, and would thus be able to recall more information from memory. In contrast, those with a more detailed or piecemeal processing style (low central coherence index for copy) would have difficulty remembering the shape after 20 minutes, given the complex figure was initially processed as fragments that were not integrated into a global context. Effectively, this hypothesis is testing whether drawing strategy and therefore underlying processing style has an effect on memory. This was indeed the case across the whole sample. Therefore regardless of the presence of an ED, a piecemeal approach to task completion can serve to overload ones memory, resulting in poorer visual memory retention. In contrast, being able to think in broader concepts can provide context for remembering, perhaps prompting recall of the details throughout the remembering process.

6.3.2.3   *Analysis 3: Lifetime Phenotypic split*

Coherence results analysed with the lifetime phenotypic split revealed largely the same picture as that reported in the classical AN/BN comparison. Those with ANR replicated the same pattern of results found for the combined AN group, in that a bias toward detail was found across both the ROCF (low central coherence index, with a high accuracy score) and the GEFT (faster and less time out errors). This pattern suggests superiority with detail in the 'pure' AN group, where global integration remains intact. Also, like the findings of analysis 2, findings for the pure BN group replicated the same pattern of results found for the combined BN group, in that both ROCF accuracy and coherence index scores were significantly lower than that of HC (moderated effect size), where no difference is found for the GEFT. This again suggests poor global integration in the pure BN group, with no evidence of a superiority of detail focus. Results from the mixed ANBN cohort more closely approximate those of the ANR group, where some evidence for enhanced detail focus is found on the GEFT where less time-out errors are observed. Although a trend is seen toward a faster median time for the ANBN group, this falls short of statistical significance with a small effect size. High accuracy scores and a low coherence index further suggest detail focus with intact global processing in the mixed cohort. The current results extend findings from the traditional AN/BN comparison by illustrating that those with lifetime AN (including those that transition to current BN) display a persistent superiority with detailed processing. In comparison, detail focus does not seem to present in those with pure BN, who

display a pattern of results more consistent with poor global integration. This differing pattern of results across diagnostic categories supports the current conceptualisation of AN and BN being distinct illnesses, where those that transition between the two seem more closely aligned with the neurocognitive profile of AN.

6.3.2.4   *Analysis 4: Extreme scores*

This analysis further investigated the results in terms of the presence of weak coherence in ED and the relationship between this concept and clinical features of the illness and comorbid psychiatric disorders.

Initial analysis divided current ED participants into those with and without extreme scores (<1 SD or <15th percentile), across both coherence tasks. Results from the ROCF identified the highest proportion of extreme scores, where approximately 45-55% of women with current ED met criteria. Results from both tasks could not be collapsed given that coherence is not a unitary concept. Rather, data was split in quartiles based on HC norms, and classified into one of four quadrants based on good or poor performance across both tasks (see Figure 22). Nearly half of all ED cases (regardless of subtype) fell into the persistent detail focus quadrant, where irrespective of task demands a local processing strategy was employed. Only 11 to 22% of current ED cases showed adaptive detail focus, where their strategy switched from local to global depending on task demands (i.e. detailed on GEFT, global on ROCF). This change of strategy would suggest that these individuals have a flexible cognitive style, in that they are able to adapt their cognitive style to best fit the context of a given situation. This hypothesis will be formally assessed in a forthcoming chapter (see chapter 10). Only a small proportion of current ED cases were found to employ persistent global focus; the only quadrant where no evidence of local processing is seen. Persistent global focus was least common in women with ANR (6%), however for ANBP and BN the proportion of cases was comparable to the number of cases with adaptive detail focus. This suggests that a complete absence of local processing in ANR is relatively uncommon.

Approximately 25% of cases across diagnoses employed maladaptive detail focus, in that a local processing strategy was employed when it was disadvantageous (ROCF) but not when it was advantageous (GEFT). Initially, it seems surprising that this quartile is endorsed to such an extent. However bearing in mind the discussion of ROCF results of those with poor accuracy scores (see 6.3.2.2), this quartile may

represent those cases where poor accuracy on the ROCF means that low coherence index scores are indicative of poor planning or attention rather than a bias toward detail. The largest proportion of cases with this pattern of results were seen in those with BN (30%), the group where significantly lower accuracy scores are seen. This would explain the same individual having a slow time on the GEFT (i.e. no superiority with detail) but a low coherence index on the ROCF (indicative of superiority with detail where accuracy is intact but not when accuracy is poor). Subsequent analysis to test this hypothesis compared ROCF copy accuracy scores of those with current ED and maladaptive detail focus (n=24) and HC women. A significant difference was found (t(103)=2.51, p=0.01) where maladaptive current ED cases showed lower accuracy scores (M=27.04, SD=3.80) with a moderate effect size. (-0.58).

Regardless of a diagnosis of ANR, ANBP or BN, a similar distribution across the four quadrants is seen. This result is found despite the lack of a significant difference between BN and HC on the GEFT, from which one would assume that rates for BN across adaptive and persistent detail focus quadrants would be well behind that of AN groups. While half the number of BN cases are seen in the adaptive detail focus compared to ANR, the BN group have a higher proportion of cases in the persistent detail focus group (1.3% more than ANR), suggesting that ED subtype is not a factor of the presence of persistent detail focus. These findings indicate that 1) the majority of current ED cases (regardless of subtype) show some form of detail focus, and 2) of those with detail focus, the vast majority employ this locally biased processing style regardless of its appropriateness in a given context.

Current ED cases where a local processing bias was observed (persistent detail focus and adaptive detail focus) were selected and compared across demographic and clinical features, and comorbid psychiatric diagnoses. Women with persistent detail focus rated themselves as moderately more depressed and more anxious. They also showed significantly higher rates of social and specific phobia, and were more likely to meet criteria for multiple (2 or more) lifetime anxiety diagnoses. The impact of comorbid anxiety diagnoses on treatment outcome was discussed above (see 6.3.1.4). All of these findings indicate poorer prognostic factors for those with persistent compared to adaptive detail focus.

6.4   General Conclusions

This chapter investigated poor set-shifting and weak coherence in an ED compared to HC group split first by the presence of an ED, then further by traditional diagnostic classifications and classifications based on illness characteristics or phenotype. The aim was to investigate whether these traits fulfilled the first criteria of an endophenotype as defined by Gottesman and Gould (2003) where the candidate endophenotype is associated with the illness.

Evidence for both set-shifting and coherence meeting criteria 1 of an endophenotype in ED are presented, replicating previous results in this area (Roberts et al., 2007b; Lopez et al., 2008c). Some inconsistencies were found across set-shifting tasks, however this is to be expected given that five tasks were employed to measure set-shifting, all of which target flexibility in addition to other aspects of executive functioning (e.g. working memory & attention). Evidence for poor flexibility was most compelling in those with lifetime AN and BN behaviours. All current ED groups had significantly more 'impaired' shifting cases compared to HC, with nearly half of the ANBP group meeting this criterion. Poor prognostic factors such as longer duration of illness and the presence of self-harming behaviours were associated with impaired shifting. Due to a lack of sensitivity in the current cohort, the CatBat task will be dropped from future discussion in this thesis although results will still be presented in the relevant tables.

Findings from the ROCF and GEFT suggest enhanced local or detailed processing in lifetime AN groups. Findings for BN were more difficult to interpret, perhaps indicative of poor global integration with normal levels of local processing. However when results were collapsed across tasks and split into quartiles based on HC norms, regardless of ED subtype nearly half of all cases were found to have persistent detail focus. The small sample size of the pure BN group requires replication before conclusions can be drawn. Replication should include additional measures of impulsivity and attention in order to determine the influence of these factors on task performance. Again poor prognostic factors such as comorbid anxiety disorders were associated with persistent rather than adaptive detail focus.

6.4.1   Limitations

Some general points are of note regarding the cohort of participants used in this chapter. As all ED participants were recruited with the overall aim of this thesis

in mind, focus was placed on identifying sister pairs with ED rather than individuals. Therefore, approximately half of the 98 clinical participants used in study 1 are from families where an ED affects more than one member. This abnormal case mix may have been a factor in the results presented here. Heightened genetic risk could have either increased or decreased deficits in neurocognitive functioning, depending on whether neurocognition is a risk factor associated with or independent of genetic risk. Further research including a heritability component is required to explore this question. Secondly, the BMI of the current AN group is notably higher than usual for a clinical AN group, falling just under 18. This illustrates the diverse nature of the current sample, who were recruited largely through outpatient and community populations. While a high proportion of participants were in an acute phase of the illness (25% BMI<16.5), an equal number of participants were in some stage of recovery (21% partial remission). Such heterogeneity within the current AN sample could mask effects of illness severity on neurocognition, however no evidence for this was found in analysis 4. Finally, HC performance on the CatBat Task, Haptic Task and GEFT differed notably to previous research, minimising and in some cases eliminating the group differences between clinical and control groups. HC were not age matched to the clinical group, and while age was run as a covariate when it correlated with outcome, this method cannot fully control for effect of age. Similarly, no measure of IQ was employed in the current study from which to also match clinical and control groups. These aspects of the case mix should be taken into account when interpreting the current findings. Limitations will be discussed further in the final chapter.

7   Study 2- The endophenotype is primarily state-independent

7.1   Background

The results presented in Study 1 (chapter 6) have highlighted poor set-shifting and weak coherence as neurocognitive traits present at a higher rate in the anorexia nervosa (AN) population compared to control women. These results fulfil the basic assumption of a candidate endophenotype, in that the trait is associated with the illness. Assessing whether these neurocognitive traits are state (present only in the active phase of the illness) or trait (present regardless of the illness being active or not) is an essential next step in understanding whether a particular characteristic is as aspect of the illness itself, or an underlying characteristic/candidate endophenotype. If such a characteristic is present in the recovered phase of the illness, this is a clear indicator that the marker is not tied to illness state in that when the illness has gone, the marker still remains.

Research in recovered populations of women with AN is sparse. The available literature has largely focussed on either enduring medical complications of the illness such as persistent amenorrhea (Brambilla et al., 2003), or enduring personality characteristics (Matsunaga, Kaye, McConaha, Plotnicov, Pollice, & Rao, 2000; Klump et al., 2004), mood disorders (Holtkamp, Muller, Heussen, Remschmidt, & Herpertz-Dahlmann, 2005) and general functioning (Wentz, Gillberg, Gillberg, & Rastam, 2001). In general, these studies have found that comorbid psychiatric conditions present in the active state of the illness have persisted in a subset of women after recovery.

Three papers to date have investigated set-shifting ability in a recovered AN population, all of which indicate that difficulties in cognitive flexibility persist after recovery (Pendleton-Jones et al., 1991; Tchanturia et al., 2002; Tchanturia et al., 2004b). Studies by Tchanturia and colleagues found no difference between currently ill (mean BMI 13.3) and weight recovered AN (mean BMI 18.4) on the Trail Making Test (TMT), Brixton test, or Haptic Illusions (Tchanturia et al., 2004b). Similarly, no differences in flexibility were found between women with current AN (mean BMI 13.7) and those recovered from AN with one year normal weight and regular periods (mean BMI 20.4). Those recovered from AN made significantly more perceptual illusions on the Haptic task compared to control women (Tchanturia et al., 2002).

Unfortunately, no papers have been found on set-shifting in a recovered BN population, and indeed this chapter will present data from recovered AN only.

One paper assessing weak coherence in a mixed recovered cohort (35 recovered AN; 7 recovered BN) has been published. This study showed that a bias toward detail remains after recovery, as measured by the Rey-Osterrieth Complex Figure (ROCF), the Embedded Figures Task and the sentence completion task (Lopez et al., 2008e). Whilst not intending to measure weak coherence per se, Pendleton-Jones et al. found that superiority on the block design task (observed in their current AN sample) weakened slightly but still persisted in weight recovered AN (Pendleton-Jones et al., 1991). The present chapter (study 2) seeks to further investigate findings from both set-shifting and weak coherence in a pure AN recovered sample, thereby addressing criteria 2 for an endophenotype "the endophenotype is primarily state-independent".

The aforementioned studies have investigated the neurocognitive profile of those recovered from an eating disorder in comparison to a HC sample, where a trait hypothesis would predict a significant finding. The present chapter will address both this standard comparison in addition to directly comparing current and recovered AN groups, where the trait hypothesis would predict no significant finding. This clinical comparison is purposeful, in order to assess the second endophenotype criteria outlined by Gottesman and Gould (2003) as directly as possible. While this criterion would be most appropriately assessed in a longitudinal design following the same group of individuals as they recover from an eating disorder, this was beyond the scope of the current thesis, thus a cross-sectional analysis is presented here. Due to the lack of sensitivity of the CatBat task in Study 1 (see 6.3.1), this chapter will focus on set-shifting results from the TMT, WCST, Brixton and Haptic tasks, in addition to the two measures of coherence.

7.1.1 Hypotheses

It is hypothesised that individuals recovered from AN will display 1) no significant difference in cognitive flexibility compared to those with current AN but impaired shifting compared to HC, and 2) no difference in detail focussed processing compared to those with current AN but a more detail focussed style compared to HC.

7.2 Method

The general methodology is outlined in Chapter 5.

7.2.1   *Participants*

Participants for this study were 30 female participants recovered from AN (18 or 60% ANR; 9 or 30% ANP; 3 or 10% ANBP), 68 with current AN (35 or 51% ANR; 24 or 35% ANP; 9 or 13% ANBP), and 88 healthy controls (HC). As outlined in 5.2.2.1, 'recovered' was defined as 1 year or more of normal BMI (>18), regular menstruation, and no eating disorder behaviour (e.g. restricting or purging). Both the current AN and HC groups used here as comparison groups are the same as those presented in Study 1. It was decided that combined current ANR and ANBP groups would be employed rather than just for example the ANR group only as, despite creating a sample size disparity between groups, approximate percentages of AN subtypes were preserved across both current and recovered AN groups.

7.2.2   *Statistical Methodology*

Neuropsychological and self-report data was assessed for normality following the procedure outlined in chapter 5.6.3. Distribution of the variables followed the same pattern as Study 1, with TMT shift time, TMT ratio, Brixton task, ROCF copy accuracy and ROCF recall accuracy only showing normal distribution. All other set-shifting and coherence variables were non-normally distributed. One-way Analysis of Variance (ANOVA) will be run for each of the normal variables, with post-hoc Tukey comparisons to examine between group differences. Non-normal variables will be assessed first with a Kruskal-Wallis test for an overall group effect, followed by Mann-Whitney U tests to examine between group differences. Of the normally distributed variables, age correlated with ROCF recall accuracy only, $r(176)=-0.28$, $p<0.001$. Analysis for this variable will be run as an ANCOVA with age as a covariate.

Extreme scores analysis for the recovered AN group followed the statistical procedure outlined in study 1 (see 6.2.7.2). A score under 15% of the HC median on the GEFT and under 1 SD of the HC mean on the ROCF coherence index was categorised as being extreme. Those with two or more extreme scores across set-shifting tasks were categorised as having impaired shifting. Extreme scores on the two coherence tasks were used to allocate each participant to one of four quadrants based on optimum task strategy; adaptive detail focus, persistent detail focus, persistent global focus and maladaptive detail focus.

Table 34: Study 2 Demographic and clinical features

	Recovered AN (n=30)	Current AN (n=68)	HC (n=88)	Test Statistic		
				F	t	p-value
Age [a]	32.13 (11.64)	24.62 (7.03)	28.43 (8.47)	8.70	-	<0.001**
Years of Education [a]	17.39 (3.18)	15.64 (2.57)	16.76 (1.98)	5.96	-	<0.01**
BMI (current) [a, b]	20.76 (1.75)	17.93 (2.60)	22.07 (1.79)	73.57	-	<0.001**
BMI (lowest)	13.85 (1.90)	14.55 (4.26)	-	-	0.87	0.38
BMI (highest)	22.67 (2.75)	21.71 (3.06)	-	-	-1.48	0.14
Current Severity [eq]	5.60 (0.50)	2.41 (1.07)	-	-	-20.13	<0.001**
Age of ED Onset	16.43 (3.05)	16.74 (4.12)	-	-	0.36	0.72
Duration of Illness	7.93 (5.52)	6.67 (5.82)	-	-	-0.21	0.83

AN Anorexia Nervosa; HC Healthy Control; BMI Body mass index
[a] Significant difference between recovered and current AN
[b] Significant difference between recovered AN and HC
[eq] Equal variances not assumed (Levene's test for equality of variance <0.05)
** Comparison significant at 0.01 level

7.3   Results

7.3.1   Demographic and clinical features

Significant demographic differences are seen between current and recovered AN for age and years of education, where the recovered group more closely approximated the HC group (see Table 34). As these two groups are matched on ED age of onset and duration of the illness, it stands to reason that the recovered group must be older and thus have had more time to pursue higher education. As expected, the BMI of the recovered AN group was significantly higher than that of the current AN group. While the recovered AN group had been recovered for on average 7.12 years (SD= 7.86), they still had an average BMI significantly lower than HC. On further inspection, a moderate, positive correlation was observed in the recovered group between current BMI and length of recovery (r(30)=0.43, p=0.02) indicating that those who had been recovered for longer had a higher BMI.

7.3.1.1   Self-report clinical features

Self-report measures revealed some notable differences between the groups (see Table 35), where those recovered from AN tended to fall between scores of the current AN and HC groups. Recovered AN reported significantly lower anxiety and depression levels than the current AN group, which remained significantly higher (with a very large effect size) than HC. While current anxiety for the recovered group remained high (above the clinical cut-off of 8), current depression levels for both groups were not marked. The recovered group also reported significantly higher levels of obsessive-compulsive behaviours and perfectionism compared to HC, but lower levels than current AN. The reverse pattern was seen for self-esteem.

7.3.1.2   Self-report cognitive style

A moderately sized difference was seen between current and recovered AN on both the cognitive flexibility scale (CFS) and the thinking styles questionnaire (TSQ). As above, the recovered AN group scored between AN and HC groups, with large to very large differences seen between recovered AN and HC groups. These results indicate a degree of improved flexibility in the recovered group, however a substantial difference still remains between recovered AN and HC.

7.3.2   Comorbidity

Lifetime comorbidity diagnoses for the recovered AN group closely mirrored those of the current AN groups reported in Study 1 (see Table 36). Nearly two thirds of those recovered from AN met lifetime criteria for depression, which had largely

Table 35: Study 2 Self-report clinical features

	Recovered AN (n=28)	Current AN (n=65)	HC (n=88)	Test Statistic			Cohen's d [2]	
				F/t	KW	p-value	AN	HC
HADS anxiety	8.12 (4.39)	11.32 (4.82)	4.20 (2.32)	68.55	-	<0.001	-0.68*	1.33**
HADS depression [1]	2 (1-4)	6 (3.25-8)	1 (0-2)	-	83.06	<0.001	-1.00**	0.80**
OCI-R total [1]	12.5 (4.25-22.25)	16.00 (7.00-27.00)	5.50 (2.25-10.00)	-	40.42	<0.001	-0.30	0.31*
Rosenberg self-esteem	18.18 (4.72)	11.82 (5.77)	23.51 (3.96)	111.04	-	<0.001	1.15**	-1.28**
Frost Perfectionism	85.77 (15.07)	96.54 (16.70)	72.95 (15.22) [3]	18.24	-	<0.001	-0.66*	0.85**
CHiRP total [1]	8 (4-11)	8 (4-11)	2.5 (1-4) [3]	-	25.34	<0.001	0.01	1.61**
Y-BOCS	12.10 (11.72)	13.77 (12.95)	-	0.59	-	0.56	-0.13	-
YBC-EDS	23.04 (6.58)	24.19 (4.93)	-	0.93	-	0.36	-0.21	-
Thinking styles	19.54 (6.42)	23.93 (6.28)	14.66 (4.89)	50.72	-	<0.001	-0.69*	0.92*
Cognitive flexibility	53.57 (7.59)	47.95 (9.16)	60.34 (5.96)	50.85	-	<0.001	0.64*	-1.06**

AN Anorexia Nervosa; HC Healthy Control; HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire; YBC-EDS Yale-Brown-Cornell Eating Disorder Scale; Y-BOCS Yale-Brown Obsessive-Compulsive Scale.
[1]Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses
[2] Cohen's d effect size comparisons with recovered AN data
[3] HC data collected from a subset of participants (n=22)
* Comparison significant at 0.05 level
** Comparison significant at 0.01 level

Table 36: Study 2 Comorbid Psychiatric Diagnoses for Recovered AN

	Full	Partial	Diagnosis[1]	Diagnostic details			
				N	Severity[2]	AOO (yrs)	DOI (yrs)
Anxiety Disorders							
OCD	8 (26.7%)	6 (20.0%)	14 (46.7%)	14	4.71 (1.20)	14.07 (5.77)	11.96 (13.71)
OCPD	8 (26.7%)	1 (3.3%)	9 (30%)	-	-	-	-
Panic Disorder	7 (23.3%)	2 (6.7%)	9 (30%)	9	4.78 (0.97)	18.89 (4.14)	5.98 (2.24)
Social Phobia	6 (21.4%)	3 (10.7%)	9 (32.1%)	9	3.56 (1.59)	11.67 (6.65)	16.44 (11.13)
Specific Phobia	2 (6.7%)	3 (10%)	5 (16.7%)	5	2.80 (1.48)	9.80 (5.50)	28.60 (17.23)
PTSD	4 (13.3%)	1 (3.3%)	5 (16.6%)	5	3.40 (0.89)	21.40 (10.99)	13.50 (7.60)
GAD	3 (10%)	1 (3.3%)	4 (13.3%)	4	2.00 (0.82)	8.25 (2.63)	24.50 (12.07)
BDD	1 (3.3%)	0	1 (3.3%)	1	5.00 (-)	11.00 (-)	2.50 (-)
Mood Disorders							
MDD	17 (56.7%)	2 (6.7%)	19 (63.4%)	19	5.21 (0.86)	19.05 (4.60)	8.62 (12.00)[3]
Bipolar	0	0	0	0	-	-	-
Dysthymia	5 (16.7%)	0	5 (16.7%)	5	5.40 (0.89)	13.80 (1.79)	13.40 (7.57)
Substance Disorders							
Alcohol A/D	4 (13.4%)	0	4 (13.4%)	3	5.00 (-)	22.33 (5.13)	2.00 (1.00)
Sub. A/D	4 (13.4%)	1 (3.3%)	5 (16.7%)	5	5.00 (1.73)	20.20 (2.17)	5.50 (6.32)

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder; Alcohol A/D Alcohol abuse/dependence; Sub. A/D Substance abuse/dependence.

[1]Diagnosis indicates pooled data from full and partial (1 criterion short) diagnosis. Severity, AOO and DOI details are for pooled data.

[2]Severity rated on 6-point scale from 1 (severe) to 6 (prior history), where 5 and 6 indicate fully recovered (>1 year)

[3]Period of time over which depressive episode/s occurred.

recovered with illness recovery. An additional five participants met threshold criteria for dysthymia, meaning that only six of the 30 participants had not suffered with a mood disorder at some point. Nine women or 30% of the recovered sample had engaged in self-harming behaviours.

Just under half of the sample met for lifetime OCD, which at the time of recruitment was also largely in the recovered state according to clinical interview. Self-report OCI-R scores were under the clinical norm of 18. The most highly endorsed OCD subtype in the recovered group was ordering (6 or 43% of those with OCD), again in line with that reported in the current AN group. Checking (n=1), hording (n=2) washing (n=3) and counting (n=2) were also endorsed. Anxiety diagnoses still in the current phase despite recovery from AN were GAD (moderate severity) and specific phobia (moderate-mild severity), where average age of onset was under 10 years old and duration of illness over 20 years. Alcohol or substance disorders were present in nearly one third of the sample, however were well recovered with a relatively short illness duration.

7.3.3   Set-shifting results

See Table 37 for descriptive statistics, test statistics and effect sizes for set-shifting tasks across groups.

7.3.3.1   Trail Making Test (TMT)

No differences were found between groups on the TMT for either the raw shift time or the balanced B-A variable. A significant difference was found between groups for the number of errors made, where the recovered AN group made significantly less errors than the current AN group, with a moderate effect size. No difference in number of errors was found between recovered AN and HC groups.

7.3.3.2   Wisconsin Card Sorting Test (WCST)

Significant overall group differences were found for both number of perseverative errors and number of categories completed. No differences were observed between current and recovered AN groups on either WCST variable, where small effect sizes were observed. The recovered AN group showed a trend toward more superior performance, with fewer errors and more categories completed than the current AN group. Compared to HC women, those recovered from AN made a similar number of perseverative errors with a negligible effect size, however the upper quartile for the recovered group is notably nearly twice than of HC. Recovered

Table 37: Study 2 Set-shifting descriptive statistics

	rcAN	AN	HC	Test Statistic			Cohen's d [2]	
	(n=24)	(n=62)	(n=78)	F	KW	p-value	AN	HC
TMT shift time (B)	27.91 (6.08)	29.31 (7.41)	28.08 (6.92)	0.69	-	0.50	-0.30	-0.13
TMT B-A	8.36 (3.02)	9.16 (6.27)	8.89 (6.31)	0.16	-	0.85	-0.14	-0.09
TMT errors (shift) [1]	0 (0-0)	0 (0-1)	0 (0-1)	-	6.15	<0.05 [£]	-0.51*	0.28
WCST Perseverative errors [1]	7 (5-16.75)	9 (7-21)	7 (5.75-9)	-	10.82	<0.01	-0.37	0.04
WCST Categories completed [1]	6 (6-6)	6 (5-6)	6 (6-6)	-	22.97	<0.001	0.28	0.67**
Brixton errors	10.17 (4.03)	11.72 (3.97)	10.01 (4.21)	4.20	-	<0.01	-0.39*	0.04
Haptic perseverations [1]	11 (7.25-19.75)	16 (10-30)	13 (7-21.75)	-	5.29	0.07	0.43*	0.11
CatBat shift time (bat)	28.91 (9.16)	31.50 (10.49)	29.08 (11.02)	1.13	-	0.34	-0.26	-0.02
CatBat B-C	8.58 (7.46)	9.37 (6.22)	8.38 (7.58)	0.53	-	0.66	-0.12	0.03
CatBat errors (shift) [1]	1 (0-1)	0 (0-1)	0 (0-1)	-	2.62	0.27	0.32	0.28

rcAN Recovered Anorexia Nervosa; AN Anorexia Nervosa; HC Healthy Control; KW Kruskal-Wallis Test; TMT Trail Making Test; WCST Wisconsin Card Sorting Test
[1]Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses
[2] Cohen's d effect size comparisons with rcAN data
* Comparison significant at 0.05 level
** Comparison significant at 0.01 level
[£] Comparison no longer significant after Hochberg correction

women completed significantly fewer categories compared to HC, with a moderate to large effect size.

7.3.3.3 Brixton Task

A significant difference between groups was found for the Brixton task, where the recovered AN group made significantly less errors than the current AN group (small-moderate effect size). The number of errors made by recovered AN and HC women did not differ (p=0.87), with an effect size close to 0.

7.3.3.4 Haptic Illusion

The recovered AN and HC groups did not differ significantly in the number of illusions experienced, however the recovered group showed a trend toward less illusions than HC. A significant, moderate effect was found between current and recovered AN groups, where those recovered perceived significantly less illusions than women with current AN.

7.3.3.5 Extreme scores

The number of women recovered from AN with extreme scores and impaired shifting was calculated. Three or 10% of women showed impaired shifting (see Figure 26).

7.3.4 Coherence results

See Table 38 for descriptive statistics, test statistics, and effect sizes for coherence tasks across groups.

7.3.4.1 GEFT

A significant group difference was found for both GEFT variables. No differences were found between current and recovered AN groups, however the recovered AN group showed both a faster GEFT median time and fewer time out errors than HC, with a moderate effect size.

7.3.4.2 ROCF

Significant group differences were found across both the ROCF central coherence index and the order and style indices. The recovered AN group showed a significantly higher score on the central coherence index compared to the current AN group, with a moderate effect size. Results were identical across both order and style indices, showing a clear pattern of global processing in the recovered AN group on the ROCF. No difference was found between the recovered AN and HC groups (p=0.93 for central coherence index), indicating comparable performance.

Figure 26: Study 2 Percentage of participants by diagnostic group with impaired set-shifting ability (extreme scores on two or more tasks)
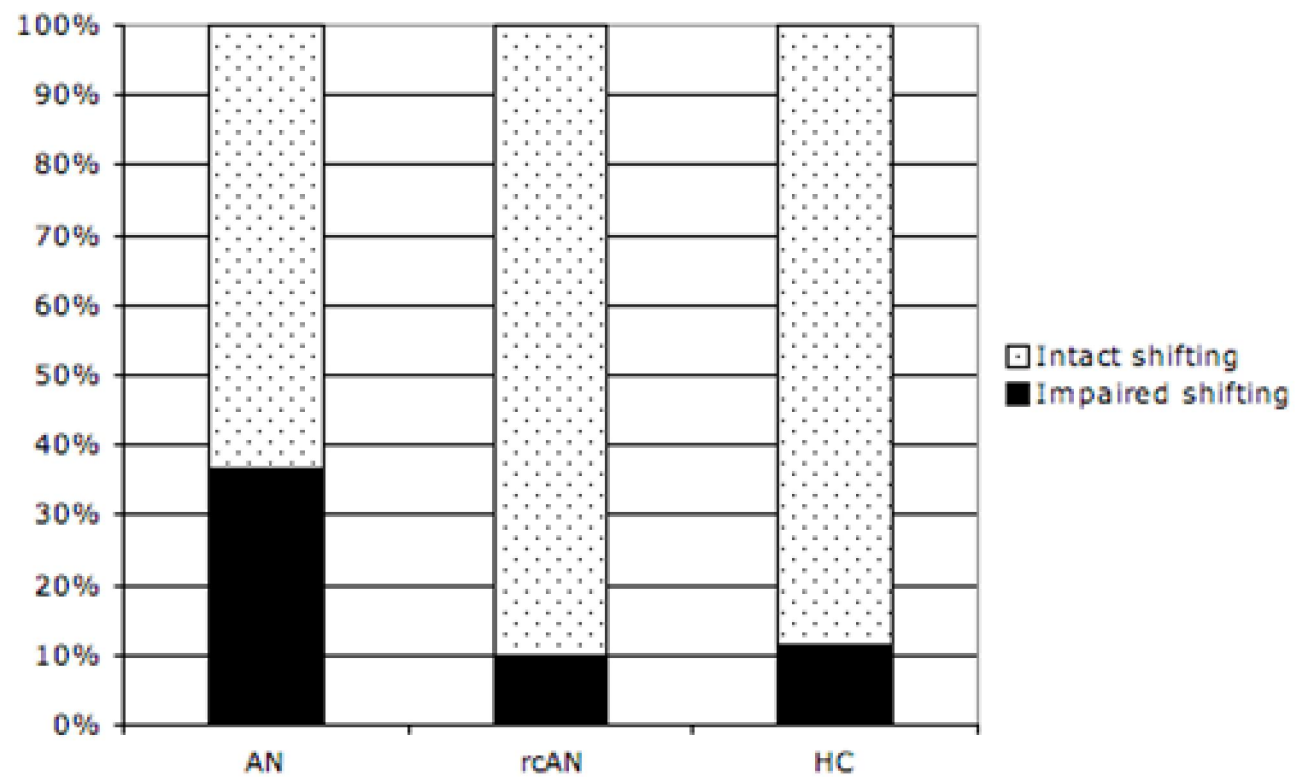
AN Anorexia Nervosa; rcAN Recovered Anorexia Nervosa; HC Healthy Control

Table 38: Study 2 Coherence descriptive statistics.

	rcAN	AN	HC	Test Statistic			Cohen's d	
	(n=27)	(n=68)	(n=81)	F	KW	p-value	AN	HC
GEFT median [1]	6.72 (5.25-7.71)	6.63 (4.63-10.74)	8.85 (5.86-15.43)	-	11.69	<0.01	0.10	0.56**
GEFT time out fails [1]	0.5 (0-2)	1 (0-2)	1 (1-2.75)	-	11.59	<0.01	0.17	0.55**
ROCF coherence index [1]	1.57 (1.40-1.7)	1.42 (1.00-1.61)	1.56 (1.41-1.56)	-	15.02	0.001	0.54*	0.06
ROCF order [1]	2.50 (2.00-2.67)	2.17 (1.81-2.5)	2.45 (2.17-2.67)	-	12.08	<0.01	0.41*	0.10
ROCF style [1]	1.67 (1.33-1.83)	1.5 (1-1.67)	1.67 (1.5-1.83)	-	15.70	<0.001	0.60**	0.17
ROCF copy accuracy	28.69 (2.40)	29.43 (3.00)	29.31 (3.92)	0.45	-	0.64	-0.26	-0.17
ROCF recall accuracy	14.54 (4.04)	15.87 (5.56)	16.58 (4.87)	6.16	-	0.001	-0.26	-0.44

rcAN Recovered Anorexia Nervosa; AN Anorexia Nervosa; HC Healthy Control; KW Kruskal-Wallis Test; GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons with HC data

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

No differences were seen between groups on copy accuracy. A significant group difference was found for recall accuracy, where age was a significant covariate (p<0.001). Post-hoc comparisons did not reach significance.

7.3.4.3 Extreme scores

Women recovered from AN were categorised into the four quartiles of coherence. The majority of cases (56.5%) showed adaptive detail focus, followed by 34.8% with persistent detail focus (see Figure 27). The remaining cases showed persistent global focus (8.7%).

7.4   Discussion

7.4.1 Set-shifting discussion

The aim of the current study was to investigate whether poor set-shifting observed in the current phase of AN (see 6.2.4.3) remains in the recovered phase, by comparing women recovered from AN with both current AN and HC women. Across the four shifting tasks, results suggest a level of recovery of set-shifting ability with recovery from the illness.

Results from the TMT (error variable only), Brixton, and Haptic tasks show that performance of the recovered AN group is significantly improved from that of current AN, and no different to that of the HC group. This suggests that those recovered from AN do not show enduring difficulties with cognitive flexibility as measured by these tasks. It should be noted that lack of a significant difference across groups on the time variables of the TMT casts doubt on the validity of its error data as a measure of set-shifting ability. As discussed in the previous chapter (see 6.3.1), the TMT is not a pure measure of set-shifting and is largely influenced by both attentional factors and individual mouse control ability. Were a higher number of errors indicative of poor set-shifting, it would be expected that longer latencies would be observed in order for the individual to cognitively shift set from their error, and adopt a new pattern of response. As this is not the case, it is possible that increased errors (but no difference in time taken) could be explained by aspects of impulsivity rather than set-shifting.

Overall group differences on the WCST were significant across both variables, however only one of four post-hoc comparisons reached significance. This is likely because the direct AN/HC comparison (presented in the previous chapter) was driving the overall effect. On comparing the recovered AN group with both

Figure 27: Study 2 Coherence strategy across tasks and diagnoses
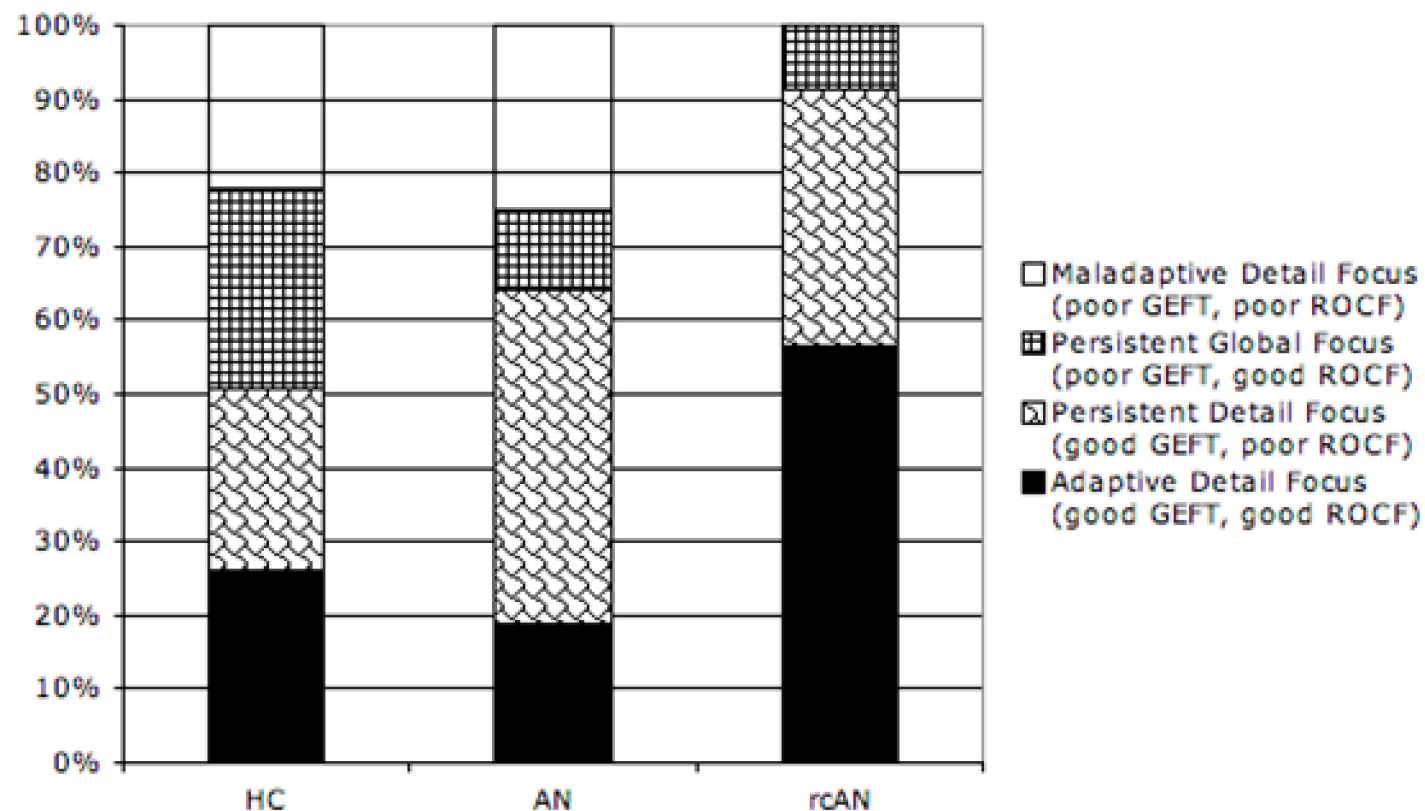
GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure; AN Anorexia Nervosa; rcAN Recovered Anorexia Nervosa; HC Healthy Control

current AN and HC women, no significant differences were found for number of perseverative errors. Examination of descriptive statistics suggests that the number of errors made by the recovered group fell between those of the current AN and HC groups (who differed significantly from each other; see 6.3.1.2). This suggests a degree or recovery, where performance of the recovered AN group trends toward that of HC but does not significantly improve compared to the current AN group. Regarding the number of categories completed, current and recovered AN groups did not differ on the WCST, however the recovered group completed significantly fewer categories than HC indicating poorer flexibility. Taken together, results from the WCST suggest a trend toward improved flexibility compared to the current AN group, however performance is still impaired compared to the HC group. This finding may indicate that the WCST is the more sensitive measure of set-shifting ability. Across all tasks, it is suggested that poor shifting improves with illness recovery to an extent, however not completely to the level of flexibility observed in the general population. In other words, cognitive rigidity may be a state independent factor that is exaggerated in the severe phase of the illness. These results provide tentative support for the hypothesis of set-shifting as a trait rather than state marker of AN, addressing criterion 2 of an endophenotype.

When collapsed across set-shifting tasks, only 10% of the recovered group exhibit 'impaired' shifting. This rate is in line with that of the HC group, and one third of that seen in women with current AN. This finding casts doubt on the state-independence of poor set-shifting, however methodological considerations must be taken into account. While descriptive statistics show a relatively consistent pattern of the recovered AN group falling between the scores of current AN and HC groups, they were significantly different to HC on the WCST only. In order to be categorised as having 'impaired' shifting, a participant must score below 1 standard deviation of the HC mean across two different set-shifting tasks (e.g.. the WCST plus one additional set-shifting task). Given the subtle nature of the group differences presented here between the recovered and current AN groups, and in the other direction recovered and HC groups, it is possible that the blunt categorisation method employed to determine intact/impaired shifting was not appropriate here.

The Brixton task has been used only once previously with a recovered AN group. Results were consistent with those presented here, where no difference was found between recovered AN and HC with a small effect size of 0.33 (Tchanturia et

al., 2004b). This illustrates that the Brixton task may be sensitive to illness state, and could act as an indicator of recovery. Despite a difference of only 1.5 errors in the current study, a moderate effect size is seen between current and recovered AN groups, illustrating a notable difference in performance across illness states on this task. While the same pattern is seen for the Haptic task, descriptive statistics indicate that in fact the recovered group make less perseverations than both current AN and the HC group, indicating not only an improvement but recovery on this task beyond HC performance. Concerns about the high score of the current HC group on this task mentioned in the previous chapter merits comparison with previous HC samples. Were the recovered AN group compared to the HC group of the aforementioned Tchanturia study, an effect size of 0.77 is seen between recovered AN and HC (in comparison to 0.01 in the current sample). If compared to the HC sample of Holliday et al. (2005), an effect size of 0.75 is seen between recovered AN and HC. Similar findings are evident when these same comparisons are made for the Haptic task. Using these HC groups as the baseline comparison, the present results would indicate that those recovered from AN do in fact show persistent cognitive and perceptual rigidity on the Brixton and Haptic tasks.

Collating these findings, it seems the Brixton task is a consistently sensitive measure of illness state, in that those recovered from the illness show normal performance on this tasks. While a trend toward improved shifting ability in recovery is seen on the WCST, overall shifting performance does not improve to the level of the general population. No previous studies of WCST performance in a recovered population can be found with which to compare this result. Given that both previous studies investigating shifting performance in recovered populations have also reported no difference overall between underweight and weight recovered individuals with AN (Pendleton-Jones et al., 1991; Tchanturia et al., 2004b), it is suggested that the hypothesis for set-shifting as a trait rather than a state marker has gained further evidence here.

7.4.2 Coherence Discussion

The current study investigated whether the weak coherence cognitive style found in those with a current diagnosis of AN (see 6.2.4.4) persisted in those recovered from AN. Results revealed mixed findings across tasks. On the GEFT, the recovered group performed the same as those with current AN and significantly faster than HC, indicating superior performance on tasks requiring detailed or local

processing. This suggests that regardless of illness state, both those with current AN and those recovered from the illness show a persistent superiority processing local stimuli. This finding replicates the results reported by Lopez (2008e), where a mixed cohort of 42 recovered AN and BN were significantly faster on the Embedded figures test (a shorter variant of the GEFT including 12 of the 18 items administered in the GEFT) compared to 42 HC. Although current and recovered AN groups were not directly compared, the recovered ED group reported by Lopez were 1.9 seconds faster (median 6.7) than current AN (Lopez et al., 2008b). This between group difference is notably larger than that observed in the current study, where AN groups were more comparable, differing by just over half a second. Despite the current HC group performing faster on the GEFT compared to the HC group of Lopez et al. (8.9 seconds compared to 12.2 seconds) the finding of superior detail focus irrespective of illness state was replicated. Persistent detail focus with weight recovered AN has also been found in a previous study using the block design task, performance on which is thought to be benefited from heightened detail focus (Pendleton-Jones et al., 1991).

In contrast, on the ROCF women recovered from AN performed the same as HC women with a central coherence index significantly higher than the current AN group. This indicates a global processing style in the recovered AN group on this task. This result is inconsistent with the one previous study using the ROCF in a recovered population, where the 42 recovered ED mentioned above showed a significantly lower central coherence index compared to HC (effect size 0.57). The Lopez et al. (2008e) recovered group consisted of nearly 20% BN and showed significantly lower copy and recall accuracy scores compared to HC. The present recovered AN group showed accuracy scores in line with the HC group. Given the consistent finding of lower accuracy in those with current BN, the subset of recovered BN participants in the Lopez study may have lowered the average accuracy score. Moreover, as lower accuracy has been found to associate with a lower coherence index (see study 1), a difference in case mix may more accurately explain the difference between studies rather than a conclusion of non-replication across these two cohorts.

In light of the seemingly contradictory findings across tasks in the recovered group, further exploration of the two tasks employed to measure weak coherence is merited. The nature of the GEFT requires the participant to "zoom in" in order to

find a local element, and therefore the optimum strategy in order to complete the task well is to employ local processing. In this scenario, the ability to focus on local or detailed elements is adaptive. Conversely, in order to copy the complex figure accurately, the optimum strategy for completing the ROCF is to be global, essentially "zooming out" to see the full context of the picture. Global perspective means that the drawing is constructed in proportion, whereas a detail focus on this task can mean that the context of the picture is lost, resulting in notable fragmentation and/or reduced accuracy. Although the optimum strategy for accuracy is to be global, the ROCF is scored (using the central coherence index) in terms of detail focus.

Consistent patterns in task strategy are seen for both the current AN and HC groups, where the AN group shows a consistent bias toward detail and the HC group show a consistent bias toward global integration. In contrast, while those recovered from AN retain the superior detail focus observed in the currently ill sample, they are capable of employing a more global strategy as dictated by contextual demands: adopting a detail focussed strategy when adaptive (GEFT) but changing to a more global strategy when preferable (ROCF). This accommodating strategy is clearly depicted when results are collapsed across tasks, and categorised into four dimensions of coherence by optimal task strategy. Over half of those recovered from AN show adaptive detail focus, in that they are able to change task strategy based on contextual demands. Such a relationship between tasks was not observed for the current AN group, where a bias toward detail was present regardless of optimum task strategy. It is possible that the trend toward a degree of improved set-shifting ability in the recovered group has allowed for switching between detailed and global strategies. These results provide support for Happe and Booth's (2007) more recent conceptualisation of the weak coherence account, where superiority with detail can but does not necessarily indicate poor global integration. It is suggested that the global integration aspect of coherence is lost in the acute phase of AN but is regained with recovery (state-dependent), whereas superiority with detail is a trait rather than state marker (i.e. state-independent).

Considering the results of both ROCF and GEFT for the present study and results reported by Lopez et al., findings lend support to the detail focussed processing style of the weak coherence hypothesis fulfilling criterion 2 for an endophenotype of AN.

7.5   General Conclusions

The findings presented here indicate that those recovered from AN retain a superiority with local processing and some persistent difficulties with set-shifting. The recovered AN group showed performance similar to HC on some set-shifting tasks, where they were significantly better at shifting than the current AN group. However on others, the recovered group performed intermediate to both current AN and HC, or were significantly more rigid than HC women. Taken together, results suggest that set-shifting recovers to a degree in those with AN, however not to the level of flexibility observed in the general population. A trend toward increased cognitive flexibility in the recovered group with longer recovery time should be noted. These findings provide tentative but not convincing support for set-shifting fulfilling the state-independence criteria of an endophenotype.

Like current AN, the recovered group showed a superior ability to process information in details or parts, however they were also capable of employing a more global strategy as dictated by contextual demands. This was evidenced by over half of those recovered from AN falling into the adaptive detail focus quadrant when collapsed across tasks. These findings provide support for the detail focus aspect of the weak coherence hypothesis fulfilling criterion 2 of an endophenotype of AN.

7.5.1   Limitations

Due to recruitment challenges, the present study was unable to include a recovered bulimia nervosa (BN) cohort. Study 1 outlined both poor set-shifting and weak coherence in both AN and BN populations. Therefore, it would be of interest to examine these traits in a cohort of women recovered from BN, to determine the state-dependence of these traits in the BN subset of eating disorders. Such an investigation would further inform these traits as endophenotypes of BN.

A disparity in sample size across groups must be noted as a limitation of the current study. A total of 30 individuals recovered from AN were included in the study. Following outlier deletion, between 24-30 individuals were included in the analysis depending on task, which was less than half the number of those with current AN and one-third the size of the HC group. This lessened statistical power of the recovered AN group may have contributed to the results presented here. Investigation of set-shifting and weak coherence with a larger recovered sample may help to clarify the results presented here.

The limitations of a cross-sectional design to address state-independence are also noted. As the recovered group had indeed managed to overcome their AN, it is possible that they represent a different cohort than those with chronic, persistent AN. The cross-sectional design employed here means that one cannot be sure whether the present recovered group even displayed for example poor set-shifting when in the current phase of their illness. While current and recovered AN groups being matched for duration of illness and lowest BMI go some way toward assurance of this limitation, little is currently known about the relationship between clinical features and neurocognitive profile. Adopting a longitudinal design to track neurocognitive profile with recovery from AN would be the best next step in assessing criterion 2 of set-shifting and weak coherence as endophenotypes of AN.

8 Study 3- Within families, the endophenotype and the illness co-segregate

8.1 Background

Large genetic studies in recent years have provided empirical evidence backing the clinical observation that eating the risk of an eating disorder (ED) increases exponentially with an affected relative (Lilenfield et al., 1998; Bulik, Sullivan, Tozzi, Furberg, Lichtenstein, & Pedersen, 2006). Study 1 found both poor set-shifting and weak coherence in women with ED compared to control women. It therefore stands to reason that family members who both have an ED would also display these traits, however it is of interest to examine this relationship explicitly. The aim of this chapter (study 3) is to address criterion 3 of an endophenotype "Within families, the endophenotype and the illness co-segregate". Both poor set-shifting and weak coherence will be examined in sister pairs both of whom have experienced an ED.

8.1.1 Hypotheses

It is hypothesised that no significant differences will be found between affected sister pairs across both 1) set-shifting tasks, and 2) coherence tasks.

8.2 Method

The general methodology is outlined in Chapter 5.

8.2.1 Participants

Participants for this study were 27 sister pairs, where both sisters had a current or lifetime diagnosis of anorexia nervosa (AN), bulimia nervosa (BN), or eating disorder not otherwise specified (EDNOS). Sisters were allocated to either 'sister 1' (AN) or 'sister 2' (ED) groups based on diagnosis, where sister 1 (where possible) had a diagnosis of current AN. The diagnostic profile for each group was as follows: Those in sister group 1 (AN) had a diagnosis of current AN (14 or 51.9% ANR; 10 or 37.0% ANBP) or recovered AN (3 or 11.1%). Those in sister group 2 (ED) had a diagnosis of current AN (6 or 22.2%) recovered AN (13 or 48.1%) current BN (4 or 14.8%, 3 with a history of AN) recovered BN (1 or 3.7%) or EDNOS (3 or 11.1%). In the case that both sisters had current AN, the sister that was first enrolled in the study was allocated to the sister 1 group.

8.2.2   Statistical methodology

Each variable was assessed for normality following the procedure outlined in 5.6.3. Normality of set-shifting variables was the same as that reported in studies 1 and 2, where Trail Making Test (TMT) raw time, TMT ratio, Brixton, CatBat raw time, and CatBat ratio were normally distributed. These variables were therefore analysed using paired-sample t-tests, with mean and standard deviations (in parentheses) displayed as descriptive statistics in the table. All other set-shifting variables were analysed using the Wilcoxon signed ranks test, with median and upper/lower quartiles (in parentheses) displayed as descriptive statistics in the table. For coherence tasks, Group Embedded Figure Test (GEFT) variables only were non-normally distributed, and therefore analysed with the Wilcoxon signed ranks test. Unlike in studies 1 and 2, all Rey-Osterrieth Complex Figure (ROCF) variables were normally distributed and therefore analysed with paired-samples t-tests. As age did not differ between the sisters, it was not assessed as a covariate.

8.3   Results

8.3.1   Demographic and Clinical Features

Due to the method of group allocation described above, the two groups differed significantly on current illness severity and current BMI (see Table 34). These clinical features did not correlate significantly with any of the neurocognitive outcome variables. The sister groups were well matched for other clinical characteristics such as duration of illness, and severity of worst illness (as illustrated by lowest ever BMI and YBC-EDS scores). The lack of difference in lowest BMI is perhaps surprising given that all the BN sisters were allocated to the ED sister group. However all but five sisters (18.5%) in this group had a lifetime AN diagnosis and therefore an AN weight at lowest BMI. The AN sister group showed a trend toward later onset of ED, however a wide variance within the group was observed.

8.3.1.1   Self-report Clinical Features

No differences on self-report measures were seen between AN and ED sister groups (see Table 40).

8.3.1.2   Self-report Cognitive Style

No differences on the thinking styles questionnaire (TSQ) or the cognitive flexibility scale (CFS) were seen between sisters.

Table 39: Study 3 Demographic and clinical features

	AN Sister (n=27)	ED Sister (n=27)	Test statistic Paired t-test	p-value
Age	25.80 (7.18)	26.08 (7.34)	-0.30	0.77
Years of Education	16.38 (2.30)	17.62 (3.84)	-1.23	0.23
BMI (current)	18.16 (2.36)	20.67 (2.46)	-4.57	<0.001**
BMI (lowest)	14.30 (1.97)	14.68 (2.56)	-0.63	0.53
BMI (highest)	21.29 (2.31)	22.35 (2.23)	-1.76	0.09
Current Severity	3.00 (1.39)	4.44 (1.45)	-3.30	<0.01**
Age of ED Onset	17.48 (5.03)	15.63 (2.91)	1.89	0.07
Duration of Illness	7.22 (3.72)	7.74 (7.11)	-0.41	0.69

AN Anorexia Nervosa; ED Eating Disorder; BMI Body mass index
** Comparison significant at 0.01 level

Table 40: Study 3 Self-report clinical features

	AN Sister (n=25)	ED Sister (n=25)	Test statistic Paired t-test	p-value	Cohen's d
HADS anxiety	10.18 (5.26)	9.27 (5.24)	0.57	0.57	0.17
HADS depression	5.18 (3.38)	3.91 (3.98)	1.10	0.28	0.34
OCI-R total	17.80 (15.80)	16.24 (12.08)	0.35	0.73	0.11
Rosenberg self-esteem	13.98 (6.16)	16.08 (6.98)	-1.02	0.32	-0.32
Frost perfectionism	92.39 (17.07)	87.85 (17.47)	0.90	0.38	0.26
CHiRP total	8.00 (3.81)	8.06 (5.59)	-0.04	0.97	-0.01
Y-BOCS	13.04 (12.15)	12.42 (12.47)	0.18	0.86	0.05
YBC-EDS	24.31 (4.87)	23.88 (6.50)	0.37	0.72	0.08
Thinking styles	23.72 (8.10)	21.88 (7.84)	0.97	0.34	0.23
Cognitive flexibility	49.92 (10.79)	50.86 (10.33)	-0.36	0.73	-0.09

AN Anorexia Nervosa; ED Eating Disorder; HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire; YBC-EDS Yale-Brown-Cornell Eating Disorder Scale; Y-BOCS Yale-Brown Obsessive-Compulsive Scale.

8.3.2   Set-shifting results

See Table 41 for descriptive statistics, test statistics, and effect sizes for set-shifting tasks by sister group. No significant differences were found between concordant sister pairs for set-shifting variables with the exception of TMT raw shift time. The ED sister group was significantly slower than the AN sister group with a large effect size.

8.3.3   Coherence results

See Table 42 for descriptive statistics, test statistics, and effect sizes for coherence tasks by sister pair. No significant differences were found between sister pairs for either the GEFT or the ROCF. The ED sister group showed a trend toward a longer median time on the GEFT, with a moderate effect size, and a trend toward poorer copy and recall accuracy on the ROCF with small to moderate effect sizes.

8.4   Discussion

Overall, these results support the hypotheses for study 3 in that sister pairs concordant for an ED showed similar performance across set-shifting and coherence tasks. Only one significant difference was found between sister pairs, where the ED sister group took longer to complete the raw shifting trial of the TMT compared to the AN sister group, with a large effect size. This finding may be attributed to the diagnostic make-up of the groups. Study 1 found that women with BN showed a significantly longer raw time on the TMT compared to both healthy controls (HC) and AN participants (see 6.2.4.3). As all BN participants in the current study were assigned to the ED sister group, it is possible that their contribution to the data resulted in the longer TMT latency. This result for the TMT differs from WCST, Brixton and Haptic task findings, where no significant differences were found (negligible to small effect sizes only). Overall, it is concluded that sister pairs performed comparably on measures of cognitive flexibility.

No differences between sister groups on coherence measures were seen. A moderate effect size was seen for GEFT median time, where the ED sister group showed a trend toward a longer median time (by 1.15 seconds) compared to the AN sister group, although both showed similar ranges. While this difference does not reach significance, results from the ED group again match closely to those previously reported in the BN group of study 1 (see 6.2.4.4) where a longer median time is found compared to the AN group. This again suggests that scores in the ED

Table 41: Study 3 Set-shifting descriptive statistics

	AN Sister (n=27)	ED Sister (n=27)	Test statistic t[3]	WSR	p-value	Cohen's d[2]
TMT total time (shift)	25.10 (5.35)	30.17 (7.41)	-3.39	-	<0.01	-0.79**
TMT ratio	1.41 (0.31)	1.54 (0.32)	-1.24	-	0.23	-0.41
TMT errors [1]	0 (0-1)	0 (0-1)	-	-1.15	0.25	-0.32
WCST Perseverative errors [1]	7.5 (7-13.5)	9 (6-17.75)	-	-0.40	0.69	-0.11
WCST Categories Completed [1]	6 (6-6)	6 (5-6)	-	-0.05	0.96	0.00
Brixton errors	10.52 (3.39)	11.04 (4.15)	-0.61	-	0.55	-0.14
Haptic perseverations [1]	16 (9-30)	13 (9-25)	-	-1.13	0.26	0.31
CatBat BAT time (shift)	30.40 (7.92)	31.59 (12.53)	-0.40	-	0.69	-0.11
CatBat ratio	1.52 (0.26)	1.42 (0.41)	0.79	-	0.44	0.29
CatBat errors [1]	0 (0-1)	1 (0-2)	-	-2.23	0.03 [£]	-0.68*

AN Anorexia Nervosa; ED Eating Disorder; WSR Wilcoxon Signed Ranks Test; TMT Trail Making Test; WCST Wisconsin Card Sorting Test

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons

[3] Paired-sample t-test

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

[£] Comparison no longer significant after Hochberg correction

Table 42: Study 3 Coherence descriptive statistics

	AN Sister (n=27)	ED Sister (n=27)	Test statistic t [3]	WSR	p-value	Cohen's d [2]
GEFT median	5.80 (4.76-9.10)	6.95 (5.23-9.00)	-	-1.92	0.06	-0.56
GEFT time out errors	1 (0-1)	1 (0-2)	-	-0.74	0.46	-0.21
ROCF coherence index[1]	1.38 (0.33)	1.43 (0.25)	-0.62	-	0.54	-0.17
ROCF order	2.17 (0.59)	2.21 (0.43)	-0.27	-	0.79	-0.08
ROCF style	1.40 (0.39)	1.47 (0.35)	-0.82	-	0.42	-0.19
ROCF copy accuracy	29.76 (2.83)	28.06 (3.17)	0.99	-	0.33	0.57
ROCF recall accuracy	17.30 (5.73)	15.24 (5.33)	1.56	-	0.13	0.37

AN Anorexia Nervosa; BN Bulimia Nervosa; WSR Wilcoxon Signed Ranks Test; GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure
[1]Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses
[2] Cohen's d effect size comparisons
[3] Paired-sample t-test

sister group may in part be due to the effect of those with BN in this group. Similarly, the ED group shows a trend toward poorer copy and recall accuracy on the ROCF, again in line with results from the BN group reported in Study 1.

8.5   General Conclusions

In sum, the findings of this study provide support for both set-shifting and coherence meeting criterion 3 of an endophenotype, in that sister pairs concordant for an ED performed similarly across tasks assessing these constructs.

8.5.1   Limitations

This study has a methodological limitation of note. The allocation of each sister into the AN or ED group resulted in an uneven spread of diagnoses across the two clinical groups. This method was chosen in order for a systematic allocation of cases to be employed. However it has meant that the potential effect of diagnosis and/or current ED severity on neuropsychological test performance has not been accounted for and may be driving some of the results. When assessed, neither current severity nor current BMI were correlated with neurocognitive outcome variables, which in some capacity suggests that current diagnosis did not affect outcome.

9   Study 4- The endophenotype is found at higher levels in unaffected relatives than the general population (AN)

9.1   Background

Perhaps the most powerful indicator of an endophenotype is the presence of the potential marker in unaffected relatives of those with the illness. That is, where unaffected relatives of those with an eating disorder (ED) also present with the candidate trait. One previous study has investigated set-shifting as a potential endophenotype, by assessing set-shifting performance in 47 females with lifetime anorexia nervosa (AN) and their sisters with no history of an ED (Holliday et al., 2005). This study found that sisters both with and without AN were significantly poorer than healthy controls (HC) on two set-shifting tasks, and did not differ significantly from each other. No significant findings emerged for the TMT and Brixton tasks. As yet, there have been no attempts to replicate this initial finding and these same traits have not been investigated in the BN population. Likewise, no research to date has looked at weak coherence in unaffected family members of those with an ED. A number of studies from the Autism literature have investigated weak coherence amongst family members. It has been found that parents (Bolte & Poustka, 2006) and particularly fathers (Happe et al., 2001) show a bias toward detail using tasks such as the Embedded Figure Test and Block Design Task.

The aim of the current chapter (study 4) is to investigate both set-shifting and weak coherence in pairs of sisters discordant for ED, that is where one sister has AN or bulimia nervosa (BN) and the other has no history of an ED. Unaffected sisters are a useful relative group to study in this context, given that they are the same gender, share a similar developmental environment and are close in age. This makes them preferable to parents as a comparison.

Focus in this chapter will be placed on the performance of unaffected sisters compared to their ED sister and HC. This investigation will address criterion 4 of an endophenotype "The endophenotype found in affected family members is found in non-affected family members at a higher rate than in the general population".

9.1.1   Hypotheses

It is hypothesised that 1) unaffected sisters of those with an ED (UA-ED) will show impaired set-shifting, in that their performance will be significantly worse (more rigid) than HC, 2) Unaffected sisters of those with AN (UA-AN) will show a

bias toward detail compared to HC, and 3) unaffected sisters of those with BN (UA-BN) will show no bias toward detail but rather poor global integration. Hypotheses are based on the findings of Study 1 (see chapter 6). Each analysis will also investigate the direct relationship between discordant sister pairs.

9.2 Method and Results

The general methodology is outlined in Chapter 5. This combined method and results section will present the findings split into three sections. Each section will assess neuropsychological performance across different sister pair groupings; 1) an assessment of all sister pairs discordant for an ED compared to HC, 2) sister pairs discordant for AN compared to HC, and 3) sister pairs discordant for BN compared to HC.

9.2.1 Method Analysis 1: ED discordant sister pairs

9.2.1.1 Participants

Participants were 50 sister pairs, where one sister had a current ED diagnosis and the other sister had no history of an ED. Thirty sister pairs were discordant for AN (13 ANR; 16 ANBP) with one AN sister recovered from the illness. Twenty sister pairs were discordant for BN, where eight (40%) of the BN sisters had a history of AN. ED sisters were also included in the current ED sample in study 1.

9.2.1.2 Statistical methodology

In line with the hypotheses, two difference analyses are planned; independent-samples t-tests to compare unaffected ED sisters with HC, and paired-samples t-tests to compare UA-ED sisters and ED sisters. Each variable was therefore assessed twice for normality following the procedure outlined in 5.6.3. For the UA-ED/HC comparison, most neuropsychological variables were normally distributed; TMT shift time; TMT B-A; WCST perseverative errors; Brixton task; CatBat Bat time; CatBat B-C; Haptic task, ROCF copy and recall accuracy. All other variables were not normal, and were therefore analysed with Mann-Whitney U tests. In the tables, descriptive statistics for normally distributed variables are presented as means with standard deviations in parentheses, while non normal variables are presented as medians with the inter quartile range in parentheses. For self-report variables, all but the HADS depression, OCI-R total scale score, CHiRP total scale score, YBC-EDS and the Y-BOCS were normally distributed.

For the UA-ED/ED sister comparison, normality was largely the same across variables, with the exception of WCST perseverative errors and the Haptic task. As means with standard deviations are presented as descriptive statistics in the table for the UA-ED group on these variables (given their assessment as parametric with the UA-ED/HC comparison), median scores with quartiles are presented as footnotes for the UA-ED sister group for these two tasks. Descriptive statistics for the ED sister group are presented as medians with quartiles for WCST and Haptic tasks in the table. These non-normal variables were analysed with Wilcoxon Signed Ranks Tests, as the data needed to be paired by family. All normally distributed variables for the direct sister pair comparison were analysed with paired-samples t-tests. Self-report variables were distributed as above. No covariates are employed, as groups did not differ significantly on age.

9.2.2   Results Analysis 1: ED discordant sister pairs
9.2.2.1   Demographic and clinical features

See Table 43 for descriptive statistics of demographic and clinical features. Unaffected ED sisters did not differ compared to either HC or their ED sisters on age or years of education. While they were matched to HC on current BMI, UA-ED sisters had a significantly higher BMI than their ED sisters. Likewise, lowest ever BMI was significantly lower in the ED sisters, however highest ever BMI did not differ. ED sisters were comparable in clinical features to the full ED group reported on in study 1 (see 6.2.2.1), where current illness severity was on average moderate (mean severity 2.2), and onset of the illness was on average at 17 years with a range of 3.5 years. Duration of illness was also similar at 8 years, with a range of 5 years.

Self-report Clinical Features: On self-report measures, UA-ED sisters reported significantly higher levels of anxiety and depression and significantly lower levels of self-esteem compared to the HC group (see Table 44). ED sisters reported higher levels of anxiety and depression compared to their UA-ED sister, along with higher current and childhood obsessive-compulsive traits. ED sisters were lower on self-esteem than their UA-ED sister, and showed higher levels of perfectionism. Like self-esteem, for self-report flexibility and thinking style UA-ED sisters reported being both significantly less flexible compared to HC, and significantly more flexible compared to their ED sister. ED sisters were significantly higher on both the YBC-EDS and the Y-BOCS total scores than UA-ED sisters.

Table 43: Study 4.1 Demographic and clinical features for ED discordant sister pairs

	UA-ED sister (n=50)	HC (n=88)	ED sister (n=50)	Test Statistics (UA-ED/HC)		Test Statistics (UA-ED/ED)	
				t	p-value	t	p-value
Age	25.63 (7.60)	28.43 (8.47)	25.26 (6.19)	1.92	0.06	0.50	0.62
Years of Education	16.03 (2.33)	16.76 (1.98)	15.70 (2.93)	1.87	0.06	0.70	0.49
BMI (current)	22.30 (2.27)	22.07 (1.79)	19.07 (3.28)	0.64	0.52	5.69	<0.001**
BMI (lowest)	20.18 (2.02)	-	15.79 (5.13)	-		5.37	<0.001**.
BMI (highest)	23.67 (2.81)	-	23.01 (3.53)	-		1.39	0.17
Current Severity	-	-	2.22 (1.20)	-		-	
Age of ED Onset	-	-	16.98 (3.51)	-		-	
Duration of Illness	-	-	8.25 (5.31)	-		-	

UA-ED Unaffected Eating Disorder; HC Healthy Control; ED Eating Disorder; BMI Body mass index
** Comparison significant at 0.01 level

Table 44: Study 4.1 Self-report clinical features for ED discordant sister pairs

	UA-ED sister (n=48)	HC (n=86)	ED sister (n=49)	Test statistics (UA-ED/HC)				Test statistics (UA-ED/ED sisters)			
				t	MW	p	Cohen's d[2]	t	z (WSR)	p	Cohen's d[2]
HADS anxiety	6.60 (3.67)	4.20 (2.32)	12.74 (4.39)	4.11	-	<0.001	0.84**	7.30	-	<0.001	-1.52**
HADS depression [1]	2 (1-4)	1 (0-2)	6 (3-11.75)	-	1260.5	<0.001	0.71**	-	-4.97	<0.001	-1.17**
OCI-R total [1]	5 (2-11) [4]	5.5 (2.25-10)	20 (13-29.75)	-	2065.0	0.83	0.04	-	-4.68	<0.001	-1.87**
Rosenberg self-esteem	20.22 (5.31)	23.51 (3.96)	10.85 (5.44)	3.78	-	<0.001	-0.73**	7.88	-	<0.001	1.74**
Frost perfectionism	73.09 (16.14)	72.95 (15.22) [3]	99.76 (14.09)	0.03	-	0.97	0.01	8.47	-	<0.001	-1.76**
CHiRP total	3.11 (2.62)	2.86 (2.78) [3]	7.32 (4.33)	0.34	-	0.73	0.09	-	-3.14	<0.01	1.17**
Y-BOCS [1]	0 (0-3)	-	15 (0-24)	-				-	-4.12	<0.001	-1.53**
YBC-EDS	1.57 (2.68)	-	23.87 (6.24)	-				23.32	-	<0.001	-4.63**
Thinking styles	19.87 (6.48)	14.66 (4.89)	25.34 (6.89)	5.30	-	<0.001	0.95**	4.44	-	<0.001	-0.82**
Cognitive flexibility	55.69 (7.69)	60.34 (5.96)	47.03 (9.00)	3.93	-	<0.001	-0.70**	4.74	-	<0.001	1.03**

UA-ED Unaffected Eating Disorder; HC Healthy Control; ED Eating Disorder; MW Mann-Whitney U Test; WSR WIlcoxon Signed Ranks Test; HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire; YBC-EDS Yale-Brown-Cornell Eating Disorder Scale; Y-BOCS Yale-Brown Obsessive-Compulsive Scale.

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons with UA-ED sister data

[3] HC data collected from a subset of participants (n=22)

[4] UA-ED CHIRP median 3 (1-4)

** Comparison significant at 0.01 level

Self-report Cognitive Style: Large differences were found between UA-ED sisters and both HC and ED sisters on the cognitive flexibility scale (CFS) and the thinking styles questionnaire (TSQ). On both scales, the scores of UA-ED sisters fell between those of HC and ED sisters.

9.2.2.2   Set-shifting results

Descriptive statistics, data analysis and effect size results for set-shifting variables (UA-ED sisters, ED sisters, HC) are presented in Table 45.

TMT: No significant differences were found between UA-ED sisters and HC, or UA-ED sisters and their ED sisters on the TMT task. Negligible to small effect sizes were seen across all comparisons. A trend was seen across both the raw and balanced TMT variables, where descriptive statistics of the UA-ED sisters fell consistently between those of the HC and ED sister groups. A moderate effect size was seen between sisters on the number of errors made, where UA-ED sisters showed a trend toward less errors than their ED sisters.

WCST: A significant difference was found for perseverative errors, where UA-ED sisters made significantly more errors than HC (moderate effect size). No difference between sisters on perseverative errors was found. However for categories completed, UA-ED sisters and HC performed comparably, with a negligible effect size. A significant difference was found between sisters with a large effect size, where UA-ED sisters completed significantly more categories than their ED sisters.

Brixton task: No significant differences were observed for either comparison on the Brixton task. The UA-ED sister group showed the highest number of errors, however this difference of one error compared to the HC group gave only a small effect size.

Haptic task: No differences in the number of illusions were found between UA-ED sisters and HC, with a negligible effect size. A significant difference was found between sisters, where UA-ED sisters made significantly less perseverations (large effect size).

9.2.2.3   Coherence results

Descriptive statistics, data analysis and effect size results for coherence variables (UA-ED sisters, ED sisters, HC) are presented in Table 46.

GEFT: No significant differences were found for either comparison across both median time and time-out error variables.

ROCF: A significant difference with a very large effect size was found

Table 45: Study 4.1 Set-shifting descriptive statistics for ED discordant sister pairs

	UA-ED sister	HC	ED sisters	Test Statistics (UA-ED/HC)				Test statistics (UA-ED/ED)			
	(n=50)	(n=88)	(n=50)	t	MW	p	Cohen's d	t	WSR	p	Cohen's d
TMT total time (shift)	29.33 (8.49)	28.08 (6.92)	30.88 (8.00)	0.91	-	0.37	0.17	0.99	-	0.33	0.19
TMT B-A	9.65 (5.80)	8.89 (6.31)	10.44 (6.27)	0.67	-	0.50	0.12	0.89	-	0.38	0.17
TMT errors [1]	0 (0-1)	0 (0-1)	1 (0-1.5)	-	1873.5	0.74	0.06	-	-1.60	0.11	0.48
WCST Perseverative errors	9.05 (3.42) [a]	7.74 (2.96)	9 (6-18.5) [1]	2.18	-	0.03 [£]	0.42*	-	-1.12	0.26	0.37
WCST Categories completed [1]	6 (6-6)	6 (6-6)	6 (5.5-6)	-	1617.5	0.68	0.08	-	-2.40	0.02 [£]	0.85*
Brixton errors	11.00 (4.03)	10.01 (4.21)	10.76 (3.99)	1.34	-	0.18	0.24	0.35	-	0.73	0.08
Haptic perseverations	14.04 (9.39) [b]	15.20 (9.41)	11 (18-30) [1]	-0.69	-	0.49	-0.12	-	-2.71	<0.01	0.84**
CatBat BAT time (shift)	32.77 (9.81)	29.08 (11.02)	30.14 (8.96)	1.84	-	0.07	0.35	1.53	-	0.13	0.36
CatBat B-C	8.61 (7.71)	8.38 (7.58)	9.27 (7.37)	0.16	-	0.88	0.03	0.37	-	0.72	0.08
CatBat errors [1]	1 (0-1)	0 (0-1)	0 (0-1)	-	1370.5	0.02 [£]	0.42*	-	-0.94	0.35	0.30

UA-ED Unaffected Eating Disorder; HC Healthy Control; ED Eating Disorder; MW Mann-Whitney U Test; WSR Wilcoxon Signed Ranks Test; TMT Trail Making Test; WCST Wisconsin Card Sorting Test

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[a] UA-ED sister WCST perseverative errors median 8 (6-11.25)

[b] UA-ED sister Haptic perseverations median 13.5 (6-21.75)

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

[£] Comparison no longer significant after Hochberg correction

Table 46: Study 4.1 Coherence descriptive statistics for ED discordant sister pairs

	UA-ED sisters (n=50)	HC (n-88)	ED sisters (n=50)	Test Statistics (UA-ED/HC)				Test Statistics (UA-ED/ED)			
				t	MW	p	Cohen's d	t	WSR	p	Cohen's d
GEFT median [1]	9.45 (5.27-13.45)	8.85 (5.86-15.43)	6.38 (4.63-10.07)	-	1805.0	0.42	0.17	-	-1.00	0.32	0.22
GEFT time out errors [1]	2 (0-3)	1 (1-2.75)	1 (0-2)	-	1900.0	0.72	0.06	-	-1.11	0.27	0.24
ROCF coherence index [1]	1.25 (0.78-1.43)	1.56 (1.41-1.69)	1.32 (0.91-1.62)	-	786.0	<0.001	1.17**	-	-0.97	0.33	0.14
ROCF order [1]	1.83 (1.33-2.33)	2.45 (2.17-2.67)	2 (1.5-2.5)	-	935.0	<0.001	1.00**	-	-0.83	0.41	0.12
ROCF style [1]	1.17 (0.67-1.55)	1.67 (1.5-1.83)	1.33 (0.83-1.67)	-	915.0	<0.001	1.10**	-	-0.73	0.47	0.11
ROCF copy accuracy	28.11 (4.46)	29.31 (3.92)	27.61 (3.71)	-1.55	-	0.13	-0.29	0.87	-	0.39	0.12
ROCF recall accuracy	13.76 (5.35)	16.58 (4.87)	14.35 (5.69)	-3.09	-	<0.01	-0.56**	0.61	-	0.54	-0.11

UA-ED Unaffected Eating Disorder; HC Healthy Control; ED Eating Disorder; MW Mann-Whitney U Test; WSR Wilcoxon Signed Ranks Test; GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure

[1]Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

** Comparison significant at 0.01 level

between UA-ED sisters and HC, where unaffected sisters had a significantly lower central coherence index (in addition to order and style indices). No differences were found between sisters on either coherence or accuracy variables. UA-ED sisters showed significantly poorer recall accuracy than HC, with a moderate effect size.

9.2.3   Method Analysis 2: AN discordant sister pairs

9.2.3.1   Participants

Participants for this study were 30 sister pairs discordant for AN (one sister with AN and the other with no ED history), along with the 88 HC reported in Study 1. For one of the sister pairs discordant for AN, the AN sister was recovered. All other cases were in a current phase of the illness.

9.2.3.2   Statistical methodology

Each comparison for each outcome variable was assessed for normality. Normally distributed variables did not differ dramatically from those reported in the UA-ED comparison (see 9.2.1.2) and were as follows: TMT raw time, TMT B-A, Brixton task, CatBat raw time, CatBat B-C, ROCF copy and recall accuracy. All other variables were not normally distributed. For self-report variables, all but HADS depression, OCI-R total, CHIRP total, Y-BOCS and YBC-EDS were normally distributed. Normality of variables was the same across both unaffected AN (UA-AN)/HC and UA-AN/ED sister comparisons. Age differed significantly between UA-AN sisters and HC, but correlated only with ROCF recall accuracy (r(111)=-0.27, p<0.01) therefore will be run as a covariate for this variable only.

9.2.4   Results Analysis 2: AN discordant sister pairs

9.2.4.1   Demographic and Clinical Features

No differences between AN sisters and their unaffected sisters were found for years of education (see Table 47). While the ANOVA for age was significant overall, no post-hoc tests were significant. Additionally a t-test between the sister pairs did not reach significance (t(58)=-0.09, p=0.93) indicating age was also constant across AN and unaffected sisters. As expected, AN sisters had significantly lower current and lowest ever BMI's, and scored significantly higher on the YBC-EDS than their unaffected sisters.

Self-report Clinical Features: On the self-report measures, unaffected sisters had scores more similar to the HC group rather than their AN sisters (see Table 48).

Table 47: Study 4.2 Demographic and Clinical Features for AN discordant sister pairs

	UA-AN sister	HC	AN sister	Test Statistic (UA-AN/HC)		Test Statistics (UA-AN/AN sister)	
	(n=30)	(n=88)	(n=30)	t	p	t	p
Age	24.23 (6.44)	28.43 (8.47)	24.10 (5.52)	2.48	0.02*	-0.16	0.87
Years of Education	15.82 (2.65)	16.76 (1.98)	15.45 (2.88)	1.76	0.09	-0.77	0.45
BMI (current)	22.27 (2.39)	22.07 (1.79)	17.45 (2.26)	-0.49	0.62	-7.52	<0.001**
BMI (lowest)	20.29 (2.14)	-	14.69 (6.10)	-		-4.55	<0.001**
BMI (highest)	23.43 (3.03)	-	21.82 (3.68)	-		-2.44	0.02*
Current Severity	-	-	2.17 (1.18)	-			
Age of ED Onset	-	-	16.67 (3.01)	-			
Duration of Illness	-	-	7.40 (4.40)	-			

UA-AN Unaffected Anorexia Nervosa; HC Healthy Control; AN Anorexia Nervosa; BMI Body mass index
* Comparison significant at 0.05 level
** Comparison significant at 0.01 level

Table 48: Study 4.2 Self-report clinical features for AN discordant sister pairs

	UA-AN sister (n=30)	HC (n=86)	AN sister (n=30)	Test statistics (UA-AN/HC)				Test statistics (UA-AN/AN sister)			
				t	MW	p	Cohen's d [2]	t	z (WSR)	p	Cohen's d [2]
HADS anxiety	5.99 (3.40)	4.20 (2.32)	11.91 (4.19)	-2.69	-	0.01	0.68*	6.08	-	<0.001	-1.55**
HADS depression [1]	1 (0-2)	1 (0-2)	5.5 (3-8.75)	-	1020.5	0.07	0.34	-	3.93	<0.001	2.24*
OCI-R total [1]	5 (1-8)	5.5 (2.25-10)	17 (10.5-28.75)	-	1227.5	0.57	0.09	-	3.74	<0.001	2.01**
Rosenberg self-esteem	20.62 (5.31)	23.51 (3.96)	10.98 (5.02)	2.74	-	<0.01	-0.65**	-6.48	-	<0.001	1.87**
Frost perfectionism	71.21 (14.31)	72.95 (15.22) [3]	101.97 (13.41)	0.42	-	0.68	-0.12	8.86	-	<0.001	-2.22**
CHiRP total [1]	3 (1.5-3)	2.5 (1-4) [3]	8 (5-11)	-	217.0	0.73	0.11	-	2.60	<0.01	2.08**
Y-BOCS [1]	0 (0-2.5)	-	18.50 (2.25-24.75)	-				-	-3.44	0.001	1.86**
YBC-EDS [1]	0 (0-1.75)	-	25 (19-28)	-				-	-4.54	<0.001	4.04**
Thinking styles	18.82 (5.88)	14.66 (4.89)	23.24 (5.27)	-3.81	-	<0.001	0.81**	3.19	-	<0.01	-0.79**
Cognitive flexibility	56.74 (6.35)	60.34 (5.96)	47.65 (7.97)	2.81	-	<0.01	-0.59**	-4.56	-	<0.001	1.26**

UA-AN Unaffected Anorexia Nervosa; HC Healthy Control; AN Anorexia Nervosa; MW Mann-Whitney U Test; WSR Wilcoxon Signed Ranks Test; HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire; YBC-EDS Yale-Brown-Cornell Eating Disorder Scale; Y-BOCS Yale-Brown Obsessive-Compulsive Scale.

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons with UA-AN sister data

[3] HC data collected from a subset of participants (n=22)

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

Low levels of depression and OCD were reported in the unaffected sister group, and while current anxiety ratings were significantly higher than controls, the mean score remained two points below the clinical cut-off and significantly lower than their AN sisters. Healthy sisters reported significantly higher self-esteem than their AN sisters, however this remained significantly lower than the self-esteem of the HC group. Current and childhood (ChiRP) perfectionism were significantly lower for unaffected compared to affected sisters.

Self-report cognitive style: UA-AN sisters scores were medial to those of their AN sisters and HC across both the TSQ and CFS. All comparisons were significant, with moderate to very large effect sizes.

9.2.4.2   Comorbidity

Unaffected AN sisters showed elevated levels of pathology in terms of depression and anxiety disorder diagnoses (see Table 49). One third of UA-AN sisters received a lifetime depression diagnosis, however at the time of testing all but two were recovered. Like their AN sisters, both specific and social phobia were highly endorsed by UA-AN sisters, where a long duration of illness was seen. Current severity ratings were on average mild, indicating that these phobias were still active in the vast majority of cases. A diagnosis of OCD was rare, with only two unaffected sisters receiving a partial diagnosis. A small number of UA-AN sisters met criteria for substance misuse disorders, with the vast majority currently recovered.

9.2.4.3   Set-shifting results

Descriptive statistics, data analysis and effect size results for set-shifting variables (UA-AN sisters, AN sisters, HC) are presented in Table 50.

TMT: No significant differences were found between unaffected sisters and either HC or AN sisters on any of the TMT variables.

WCST: A significant difference was found for number of perseverative errors, where UA-AN sisters made significantly more errors than HC, with a moderate effect size. Sisters did not differ on the number of errors made. However for number of categories completed, UA-AN sisters did not differ to HC but completed significantly more categories than their AN sisters, with a moderate/large effect size.

Brixton task: No significant differences were found on the number of errors made on Brixton task, with negligible to small effect sizes seen.

Table 49: Study 4.2 Comorbid Psychiatric Diagnoses for unaffected AN sisters (n=30)

| | Full | Partial | Diagnosis[1] | Diagnostic details | | | |
				N	Severity[2]	AOO (yrs)	DOI (yrs)
Anxiety Disorders							
OCD	0	2 (6.7%)	2 (6.7%)	2	5.00 (0.00)	14.00 (1.41)	3.25 (0.35)
OCPD	2 (6.7%)	0	2 (6.7%)	-			
Panic Disorder	3 (10.3%)	2 (6.8%)	5 (17.1%)	5	5.20 (1.10)	18.20 (4.66)	6.20 (7.56)
Social Phobia	4 (13.8%)	6 (20.7%)	10 (24.5%)	8	3.25 (1.39)	7.50 (2.78)	14.38 (4.90)
Specific Phobia	1 (3.4%)	7 (24.1%)	8 (27.5%)	8	2.50 (0.76)	10.25 (8.31)	13.94 (8.06)
PTSD	2 (6.9%)	1 (3.4%)	3 (10.3%)	2	4.50 (0.71)	24.00 (4.24)	3.00 (2.83)
GAD	0	0	0	-			
BDD	0	0	0	-			
Mood Disorders							
MDD	8 (26.7%)	2 (6.7%)	10 (33.4%)	10	5.44 (0.88)	19.10 (5.22)	2.55 (4.81)[3]
Bipolar	1 (3.3%)	0	1 (3.3%)	1	3.00 (-)	20.00 (-)	13.00 (-)
Dysthymia	0	0	0	-			
Substance Disorders							
Alcohol A/D	4 (13.3%)	0	4 (13.3%)	3	5.33 (0.58)	17.67 (0.58)	2.23 (2.38)
Sub. A/D	3 (10%)	0	3 (10%)	2	5.00 (1.41)	15.00 (-)	0.25 (-)

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder; Alcohol A/D Alcohol abuse/dependence; Sub. A/D Substance abuse/dependence.

[1]Diagnosis indicates pooled data from full and partial (1 criterion short) diagnosis. Severity, AOO and DOI details are for pooled data.

[2]Severity rated on 6-point scale from 1 (severe) to 6 (prior history), where 5 and 6 indicate fully recovered (>1 year)

[3]Period of time over which depressive episode/s occurred.

Table 50: Study 4.2 Set-shifting descriptive statistics for AN discordant pairs

| | UA-AN sister | HC | AN sisters | Test Statistics (UA-AN/HC) | | | | Test Statistics (UA-AN/AN) | | | |
	(n=30)	(n=88)	(n=30)	t	MW	p	Cohen's d	t	WSR	p	Cohen's d
TMT total time (shift)	27.43 (8.58)	28.08 (6.92)	29.05 (6.62)	0.41	-	0.69	-0.09	0.85	-	0.40	-0.21
TMT B-A	7.63 (4.44)	8.89 (6.31)	8.83 (5.49)	1.16	-	0.25	-0.21	1.32	-	0.20	-0.24
TMT errors [1]	0 (0-1)	0 (0-1)	1 (0-1.5)	-	1086.5	0.48	0.13	-	-1.34	0.18	-0.37
WCST Perseverative errors [1]	9.5 (6-14)	7 (5.75-9)	9 (6-26)	-	683.0	0.01 [£]	0.49*	-	-0.96	0.34	-0.29
WCST Categories completed [1]	6 (6-6)	6 (6-6)	6 (4.25-6)	-	988.0	0.41	0.16	-	-2.21	0.03 [£]	0.67*
Brixton errors	10.33 (4.16)	10.01 (4.21)	11.33 (3.98)	-0.36	-	0.72	0.08	0.97	-	0.34	-0.25
Haptic perseverations [1]	12 (4.5-22)	13 (7-21.75)	15.5 (10.75-30)	-	1140.0	0.26	0.02	-	-1.97	0.05 [£]	-0.25*
CatBat BAT time (shift)	30.81 (9.01)	29.08 (11.02)	30.17 (9.81)	-0.72	-	0.48	0.16	-0.66	-	0.52	0.07
CatBat B-C	8.04 (7.86)	8.38 (7.58)	9.14 (5.65)	0.19	-	0.85	-0.04	0.59	-	0.56	-0.16
CatBat errors [1]	1 (0-1)	0 (1)	0 (0-1)	-	882.0	0.24	0.23	-	-1.09	0.28	0.32

UA-AN Unaffected Anorexia Nervosa; HC Healthy Control; AN Anorexia Nervosa; MW Mann-Whitney U Test; WSR Wilcoxon Signed Ranks Test; TMT Trail Making Test; WCST Wisconsin Card Sorting Test

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

* Comparison significant at 0.05 level

[£] Comparison no longer significant after Hochberg correction

Haptic task: No difference was found between UA-AN sisters and HC for the number of illusions on the Haptic task, with an effect size close to 0. UA-AN sisters showed a trend toward less illusions than AN sisters, with a small effect size.

9.2.4.4   Coherence results

Descriptive statistics, data analysis and effect size results for coherence variables (UA-AN sisters, AN sisters, HC) are presented in Table 51.

GEFT: The median time comparison between UA-AN sisters and HC fell on the significance level of 0.05, where UA-AN sisters showed a trend toward a faster time compared to HC. A small/moderate effect size was seen, where UA-AN sisters were 2.2 seconds faster on average than HC. No difference was found between sister pairs.

ROCF: A significant difference was found between UA-AN sisters and HC across both coherence index and order/style indices, with large to very large effect sizes. UA-AN sisters remembered significantly less than HC at 20 minute recall, with a moderate effect size, where age was a significant covariate (p<0.001). No differences were found between sisters on coherence index or copy/recall accuracy.

9.2.5   Method Analysis 3: BN discordant sister pairs

9.2.5.1   Participants

Participants were 20 sister pairs where one sister had lifetime BN and one sister no history of an ED. Eight or 40% of the BN sisters had a history of AN.

9.2.5.2   Statistical analysis

Normality was assessed for each variable across both comparisons. Variable normality was found to be the same as that reported above for the UA-AN analysis (see 9.2.3.2). One difference was present, where the UA-BN/HC comparison for the Haptic task was normally distributed.

9.2.6   Results Analysis 3: BN discordant sister pairs

9.2.6.1   Demographic and Clinical Features

Healthy BN sisters did not differ to either BN sisters of HC on age, years of education, and current BMI (see Table 52). Lowest ever BMI was significantly lower in the BN group compared to UA-BN sisters, as would be expected given the AN history in just under half of the sample. BN sisters were on average in a moderate state of the illness, and had been ill for just under 10 years.

Table 51: Study 4.2 Coherence descriptive statistics for AN discordant sister pairs

	UA-AN sisters	HC	AN sisters	Test Statistics (UA-AN/HC)				Test Statistics (UA-AN/AN)			
	(n=30)	(n-88)	(n=30)	t/F	MW	p	Cohen's d	t	WSR	p	Cohen's d
GEFT median [1]	6.65 (3.85-12.25)	8.85 (5.86-15.43)	6.00 (3.9-8.69)	-	923.0	0.05	0.37*	-	-0.53	0.59	0.14
GEFT time out errors [1]	1 (0-2)	1 (1-2.75)	0 (0-1)	-	959.5	0.08	0.34	-	-1.14	0.26	0.32
ROCF coherence index [1]	1.25 (0.72-1.43)	1.56 (1.41-1.69)	1.38 (0.88-1.65)	-	432.5	<0.001	1.14**	-	-1.49	0.14	0.39
ROCF order [1]	1.83 (1.33-2.17)	2.45 (2.17-2.67)	2.17 (1.64-2.50)	-	463.0	<0.001	1.09**	-	-1.52	0.13	0.40
ROCF style [1]	1.25 (0.63-1.54)	1.67 (1.50-1.83)	1.45 (0.79-1.67)	-	601.5	<0.001	0.85**	-	-1.14	0.25	0.30
ROCF copy accuracy	29.60 (4.04)	29.31 (3.92)	29.10 (3.20)	-0.35	-	0.73	0.07	-0.52	-	0.61	0.14
ROCF recall accuracy [2]	14.13 (5.29)	16.58 (4.87)	15.40 (6.09)	10.18	-	<0.001	-0.49*	0.84	-	0.41	-0.22

UA-AN Unaffected Anorexia Nervosa; HC Healthy Control; AN Anorexia Nervosa; MW Mann-Whitney U Test; WSR Wilcoxon Signed Ranks Test; GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure

[1]Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Age run as a covariate, therefore F statistics presented

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

Table 52: Study 4.3 Demographic and Clinical Features of BN Sister Pairs

	UA-BN sister	HC	BN sister	Test Statistics (UA-BN/HC)			Test Statistics (UA-BN/BN)	
	(n=20)	(n=88)	(n=20)	t	p		t	p
Age	27.60 (8.71)	28.43 (8.47)	27.00 (6.86)	0.39	0.69		-0.58	0.57
Years of Education	16.45 (1.70)	16.76 (1.98)	16.11 (3.04)	0.65	0.52		-0.21	0.84
BMI (current)	22.40 (2.09)	22.07 (1.79)	21.51 (3.09)	-0.72	0.47		-1.09	0.29
BMI (lowest)	20.13 (1.88)	-	17.45 (2.52)	-			-3.35	<0.01**
BMI (highest)	24.03 (2.39)	-	24.75 (2.47)	-			1.27	0.22
Current Severity	-	-	2.30 (1.26)	-				
Age of ED Onset	-	-	17.45 (4.20)	-				
Duration of Illness	-	-	9.52 (6.36)	-				

UA-BN Unaffected Bulimia Nervosa; HC Healthy Control; BN Bulimia Nervosa; BMI Body mass index
[eq] Equal variances not assumed (Levene's test for equality of variance <0.05)
** Comparison significant at 0.01 level

Self-report Clinical Features: UA-BN sisters had significantly higher levels of anxiety and depression compared to HC and significantly lower levels compared to their BN sisters (see Table 53). Likewise, self-esteem for UA-BN sisters was significantly higher than BN but significantly lower than HC. Levels of perfectionism were similar across UA-BN sisters and HC. As expected, BN sisters scored significantly higher on ED preoccupations and rituals (as measured by the YBC) compared to their unaffected sisters, whose scores were close to zero.

Self-report Cognitive Style: As reported for UA-AN sisters, UA-BN sisters scored significantly different compared to their BN sisters, and in the opposite direction to HC. Differences were large to very large. This indicates that UA-BN sisters perceived themselves as more flexible compared to their BN sisters, but less flexible than the general population (HC).

9.2.6.2 Comorbidity

With the exception of social phobia and major depressive disorder, threshold psychiatric diagnoses for unaffected BN sister group are relatively low (see Table 54). While those with a depression diagnosis were considered well recovered at the time of testing, those with social phobia were still in an active stage, with an average of 16 years of illness. While social and specific phobia diagnostic rates were comparable to that of their BN sisters, alcohol or substance abuse/dependence rates were not, with little evidence substance issues in the unaffected BN sisters. OCD diagnostic rates were also low, with only one sister meeting for threshold OCD (washing & checking subtypes).

9.2.6.3 Set-shifting results

Descriptive statistics, data analysis and effect size results for set-shifting variables (UA-BN sisters, BN sisters, HC) are presented in Table 55.

TMT: A significant difference was found for the TMT balanced B-A variable, where UA-BN sisters had a significantly longer time compared to HC with a moderate effect size. No differences were found between UA-BN and BN sisters.

WCST: No significant differences were found on the WCST when comparing UA-BN with HC and BN sisters, with negligible to small effect sizes. UA-BN sister descriptive statistics were similar to those of the HC group.

Brixton task: A significant, moderate difference was found between UA-BN sisters and HC, where unaffected sisters made more errors than HC women. Despite a moderate effect size, no significant difference was found between sister pairs.

Table 53: Study 4.3 Self-report Clinical Features for BN Sister Pairs

	UA-BN sister (n=20)	HC (n=86)	BN sister (n=20)	Test statistics (UA-BN/HC)				Test Statistics (UA-BN/BN)			
				t	MW	p	Cohen's d [2]	t	z (WSR)	p	Cohen's d [2]
HADS anxiety	7.32 (4.06)	4.20 (2.32)	14.00 (4.50)	-3.24	-	<0.01	1.15**	4.09	-	0.001	-1.56**
HADS depression [1]	3.5 (2-6)	1 (0-2)	10 (3-13)	-	256.5	<0.001	1.06**	-	-3.00	<0.01	1.99*
OCI-R total [1]	6 (3-15)	5.5 (2.25-10)	22 (16-31)	-	708.5	0.30	0.23	-	-2.68	<0.01	1.62*
Rosenberg self-esteem	19.28 (5.46)	23.51 (3.96)	10.66 (6.16)	4.00	-	<0.001	-0.99**	-4.43	-	<0.001	1.48**
Frost perfectionism	76.37 (18.24)	72.95 (15.22) [3]	96.39 (14.78)	-0.65	-	0.52	0.20	3.50	-	<0.01	-1.29**
CHIRP total [1]	3 (1-6)	2.5 (1-4) [3]	4 (2-10)	-	150.5	0.65	0.18	-	-1.63	0.10	1.56
Y-BOCS [1]	0 (0-4.75)	-	14 (0-23)	-				-	-2.28	0.02	1.23*
YBC-EDS [1]	0 (0-3.75)		26 (20-30)	-				-	-5.97	<0.001	x[4]**
Thinking styles	21.80 (7.10)	14.66 (4.89)	28.54 (7.94)	-5.38	-	<0.001	1.33**	3.05	-	<0.01	-0.90**
Cognitive flexibility	54.32 (9.23)	60.34 (5.96)	46.07 (10.57)	2.79	-	0.01	-0.90*	-2.20	-	0.04	0.83*

UA-BN Unaffected Bulimia Nervosa; HC Healthy Control; BN Bulimia Nervosa; MW Mann-Whitney U Test; WSR Wilcoxon Signed Ranks Test; HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire; YBC-EDS Yale-Brown-Cornell Eating Disorder Scale; Y-BOCS Yale-Brown Obsessive-Compulsive Scale.

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[2] Cohen's d effect size comparisons with UA-BN sister data

[3] HC data collected from a subset of participants (n=22)

[4] Value could not be calculated

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

Table 54: Study 4.3 Psychiatric diagnoses for unaffected BN sisters

	Full	Partial	Diagnosis[1]	N	Severity[2]	AOO (yrs)	DOI (yrs)
					Diagnostic details		
Anxiety Disorders							
OCD	1 (5.0%)	3 (15.0%)	4 (20.0%)	4	3.25 (1.26)	12.75 (4.57)	12.00 (10.49)
OCPD	0	1 (5.0%)	1 (5.0%)	-			
Panic Disorder	2 (10.0%)	0	2 (10.0%)	2	5.00 (1.42)	17.50 (2.12)	5.75 (6.01)
Social Phobia	4 (20.0%)	4 (20.0%)	8 (40.0%)	8	2.88 (0.99)	8.88 (4.12)	16.25 (7.76)
Specific Phobia	0	4 (20.0%)	4 (20.0%)	4	2.00 (0.82)	17.50 (7.19)	7.50 (6.14)
PTSD	1 (5.0%)	0	1 (5.0%)	1	6.00 (-)	9.00 (-)	2.00 (-)
GAD	0	0	0	-			
BDD	0	0	0	-			
Mood Disorders							
MDD	4 (20.0%)	2 (10.0%)	6 (30.0%)	6	5.33 (0.82)	21.17 (6.77)	3.42 (4.75)[3]
Bipolar	0	0	0	-			
Dysthymia	0	0	0	-			
Substance Disorders							
Alcohol A/D	2 (10.0%)	0	2(10.0%)	1	6.00 (-)	24.00 (-)	2.00 (-)
Sub. A/D	1 (5.0%)	0	1(5.0%)	1	6.00 (-)	20.00 (-)	

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder; Alcohol A/D Alcohol abuse/dependence; Sub. A/D Substance abuse/dependence.

[1]Diagnosis indicates pooled data from full and partial (1 criterion short) diagnosis. Severity, AOO and DOI details are for pooled data.

[2]Severity rated on 6-point scale from 1 (severe) to 6 (prior history), where 5 and 6 indicate fully recovered (>1 year)

[3]Period of time over which depressive episode/s occurred.

Table 55: Study 4.3 Set-shifting descriptive statistics for BN discordant sister pairs

	UA-BN sister (n=20)	HC (n=88)	BN sisters (n=20)	Test Statistics (UA-BN/HC)				Test Statistics (UA-BN/BN)			
				t	MW	p	Cohen's d	t	WSR	p	Cohen's d
TMT total time (shift)	31.48 (8.22)	28.08 (6.92)	33.62 (9.21)	-1.86	-	0.07	0.47	0.52	-	0.61	-0.25
TMT B-A	12.21 (6.68)	8.89 (6.31)	12.78 (6.71)	-2.05	-	0.04 [£]	0.52*	0.09	-	0.93	-0.09
TMT errors [1]	0 (0-1)	0 (0-1)	1 (0-1.75)	-	753.0	0.64	0.09	-	-0.77	0.44	0.25
WCST Perseverative errors [1]	7 (6-9.5)	7 (5.75-9)	9 (6-13)	-	648.5	0.89	0.03	-	-0.87	0.39	0.33
WCST Categories completed [1]	6 (6-6)	6 (6-6)	6 (6-6)	-	616.0	0.07	0.17	-	-1.00	0.32	0.39
Brixton errors	12.15 (3.62)	10.01 (4.21)	9.90 (3.95)	-2.10	-	0.04 [£]	0.52*	-1.64	-	0.12	0.59
Haptic perseverations	14.84 (9.23) [a]	15.20 (9.41)	24 (11-30) [1]	0.15	-	0.88	-0.04	-	-1.97	0.05	0.68
CatBat BAT time (shift)	35.19 (10.26)	29.08 (11.02)	30.08 (7.65)	-2.21	-	0.03 [£]	0.56*	-1.70	-	0.11	0.57
CatBat B-C	8.53 (8.34)	8.38 (7.58)	9.47 (9.72)	-0.07	-	0.94	0.02	0.04	-	0.97	-0.10
CatBat errors [1]	1 (0-1)	0 (1)	0.50 (0-1.25)	-	546.5	0.02 [£]	0.46*	-	-0.23	0.82	0.08

UA-BN Unaffected Bulimia Nervosa; HC Healthy Control; BN Bulimia Nervosa; MW Mann-Whitney U Test; WSR Wilcoxon Signed Ranks Test; TMT Trail Making Test; WCST Wisconsin Card Sorting Test

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

[a] UA-BN sister Haptic perseverations median 14 (6-21)

* Comparison significant at 0.05 level

[£] Comparison no longer significant after Hochberg correction

Haptic task: UA-BN sisters and HC scored similarly on the Haptic task, where BN sisters reported significantly more illusions than their unaffected sisters (moderate effect size).

9.2.6.4   Coherence results

Descriptive statistics, data analysis and effect size results for coherence variables (UA-BN sisters, BN sisters, HC) are presented in Table 56.

GEFT: No significant differences were found between UA-BN sisters and either HC or BN sisters on either median time or time-out errors. UA-BN sisters were slower than the HC group.

ROCF: A significant difference was found on the ROCF coherence index, where UA-BN sisters had a lower coherence index (and order/style indices) compared to HC, with moderate/large effect sizes. No differences were observed between unaffected and BN sisters. For accuracy, UA-BN sisters scored lower than HC on both copy and recall. Scores were similar to BN sisters.

9.2.7   Method Analysis 4: Extreme scores

9.2.7.1   Participants

Participants were the aforementioned 50 sister pairs discordant for an ED; 30 where one sister has AN and the other sister has no lifetime history of an ED, and 20 where one sister has BN and the other sister has not lifetime history of an ED. The 88 HC women were also included.

9.2.7.2   Statistical methods

Statistical methodology followed that outlined for extreme score analysis in study 1 (see 6.2.7.2). A cut-off score for each neuropsychological task was calculated, from which participants were categorised as having an extreme score or not across each of the four measures of set-shifting and two measures of coherence.

Set-shifting analyses: A composite score was created where participants were categorised as having 'impaired' set-shifting if they had an extreme score on two or more set-shifting tasks, or 'intact' shifting if they had one or no extreme scores. Data was analysed by investigating frequencies and by running Pearson's chi-square tests to investigate differences in the number of cases with extreme scores across diagnoses. Pearson's chi-square test was also employed to explore differences in comorbidity based on shifting ability in unaffected ED sisters. Independent-samples t-tests were used to investigate differences in demographic and clinical features

Table 56: Study 4.3 Coherence descriptive statistics for BN discordant sister pairs

	UA-BN sisters (n=20)	HC (n-88)	BN sisters (n=20)	Test Statistics (UA-BN/HC)				Test Statistics (UA-BN/BN)			
				t	MW	p	Cohen's d	t	WSR	p	Cohen's d
GEFT median [1]	9.85 (6.3-15.9)	8.85 (5.86-15.43)	8.6 (5.55-13.13)	-	710.0	0.45	-0.15	-	-0.78	0.44	0.28
GEFT time out errors [1]	2 (1-3)	1 (1-2.75)	1 (1-3)	-	646.5	0.19	-0.26	-	-0.42	0.67	0.15
ROCF coherence index [1]	1.28 (0.86-1.58)	1.56 (1.41-1.69)	1.23 (0.91-1.48)	-	394.5	<0.001	0.76**	-	-0.36	0.72	0.12
ROCF order [1]	2 (1.42-2.33)	2.45 (2.17-2.67)	1.83 (1.5-2.17)	-	477.0	<0.01	0.60**	-	-0.57	0.57	0.21
ROCF style [1]	1.19 (0.88-1.65)	1.67 (1.5-1.83)	1.25 (0.83-1.73)	-	388.0	<0.001	0.78**	-	-0.12	0.90	0.04
ROCF copy accuracy	26.18 (4.45)	29.31 (3.92)	25.38 (3.33)	3.12	-	<0.01	-0.78*	-0.75	-	0.46	0.20
ROCF recall accuracy	13.58 (5.69)	16.58 (4.87)	12.68 (4.65)	2.39	-	0.02	-0.60*	-0.03	-	0.77	0.17

UA-BN Unaffected Bulimia Nervosa; HC Healthy Control; BN Bulimia Nervosa; MW Mann-Whitney U Test; WSR Wilcoxon Signed Ranks Test; GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure

[1]Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

* Comparison significant at 0.05 level

** Comparison significant at 0.01 level

across unaffected ED sisters with impaired/intact shifting. Mann-Whitney U tests were employed for non-normal data. For categorical data, Pearson's chi-square test was used.

Coherence analyses: A composite score was created by splitting data from both tasks into quartiles based on HC results. Participants were then categorised across both tasks according to their quartile placement (lower 2 quartiles or upper 2 quartiles on each task) to create a dimensional variable. Unaffected ED sisters falling into the adaptive and persistent detail focus dimensions were further compared across demographic, clinical features and comorbidity as outlined above for intact/impaired set-shifting.

9.2.8 Results Analysis 4: Extreme scores

9.2.8.1 Set-shifting results

Frequency analysis: Table 57 outlines the percentage of each diagnostic group with impaired/intact set-shifting, in that extreme scores on two or more set-shifting tasks were present. AN and BN sisters showed the expected pattern of a higher proportion of impaired cases (33.3% and 30%, respectively) compared to HC. Unaffected BN sisters showed a similar proportion of impaired cases as their BN sisters (26.3%). However unaffected AN sisters did not with only 10% meeting criteria for impaired shifting, comparable to the HC group. Pearson's chi-square test revealed that AN discordant sisters differed significantly in the number of impaired shifting cases (p=0.03) however BN sisters did not (p=0.72).

Demographic, self-report clinical features and comorbidity: Both separate and combined comparison between unaffected AN and BN sisters with impaired/intact shifting was not feasible given the small number of cases in the impaired shifting group (AN n=3, BN n=5). Some interesting observations were made following investigation of descriptive statistics. No unaffected ED sisters with impaired shifting met diagnostic criteria for depression or received 2 or more anxiety diagnoses. However 38.1% of unaffected ED sisters with intact shifting met for depression, and 35.7% met for 2 or more anxiety diagnoses.

Table 57: Study 4.4 Percentage of participants by diagnostic group with impaired set-shifting ability (extreme scores on two or more tasks)
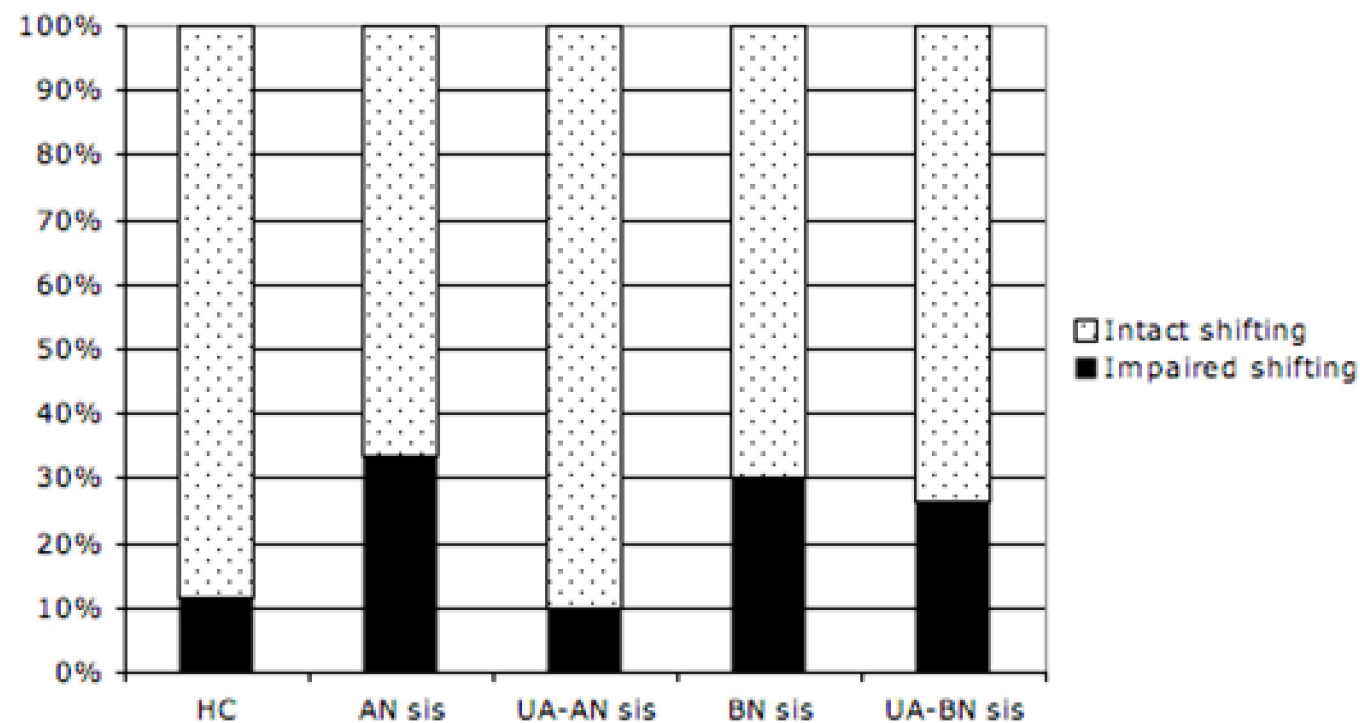
HC Healthy Control; AN sis Anorexia Nervosa sister; UA-AN sis Unaffected Anorexia Nervosa sister; BN sis Bulimia Nervosa sister; UA-BN sis Unaffected Bulimia Nervosa sister

Figure 28: Study 4.4 Weighted percentage of participants with impaired shifting by diagnostic group
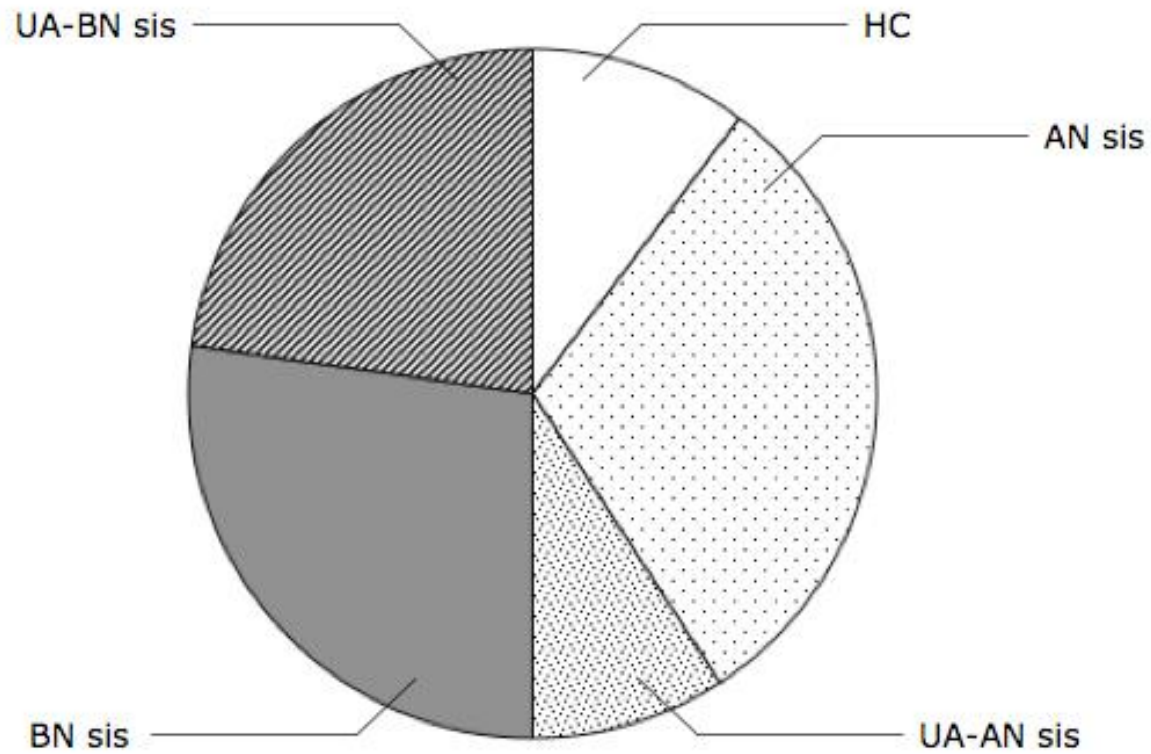
HC Healthy Control; AN sis Anorexia Nervosa sister; UA-AN sis Unaffected Anorexia Nervosa sister; BN sis Bulimia Nervosa sister; UA-BN sis Unaffected Bulimia Nervosa sister

9.2.8.2   Coherence results

Frequency analysis: Figure 29 outlines the strategy employed across both coherence tasks for affected and unaffected ED sisters and HC. Unaffected AN sisters mirrored their AN sisters in the proportion of cases employing persistent detail focus (48.3% vs. 48.1%) and persistent global focus (3.4% vs 3.7%). Few UA-AN sister cases (n=2) showed adaptive detail focus, while over 40% showed maladaptive detail focus. Unaffected BN sisters showed notably fewer cases of persistent detail focus (27.8% vs. 47.1%) and notably more cases of persistent global focus (22.2% vs. 11.8%) compared to BN sisters. High rates of maladaptive detail focus were found for both unaffected and affected BN sisters (44.4% vs. 35.3%).

Demographic, self-report clinical features and comorbidity: Unaffected ED sisters with either adaptive or persistent detail focus were selected for further analysis. As with set-shifting, given the small numbers of unaffected sisters with adaptive detail focus (AN n=2, BN n=1), comparisons between these groups was not feasible.

<div align="center">9.3   Discussion</div>

This chapter aimed to investigate set-shifting and coherence performance of the unaffected sisters of women with an eating disorder (ED) across three diagnostic groups; unaffected sisters of women with an eating disorder (ED) i.e. anorexia (AN) and bulimia nervosa (BN), then split into unaffected sisters of those with lifetime AN, and unaffected sisters of those with lifetime BN. This section will therefore follow the same format, discussing the findings from each of the three analyses first for set-shifting, then for coherence.

Overall, the hypothesis for set-shifting meeting criterion 4 of an endophenotype received modest support, in that sisters of women with a diagnosis of ED displayed a more rigid or inflexible profile on some tasks compared to HC women. When collapsed across tasks, unaffected BN sisters showed more evidence of impaired set-shifting than unaffected AN sisters. The hypothesis for weak coherence in unaffected sisters of those with ED received strong support across tasks, where sisters of women with AN and BN differed in their manifestation of weak coherence in line with the profile of their affected sisters. When data was collapsed across optimum task strategy, unaffected AN sisters showed the higher prevalence of persistent detail focus.

Figure 29: Study 4.4 Coherence strategy across both tasks by diagnostic group
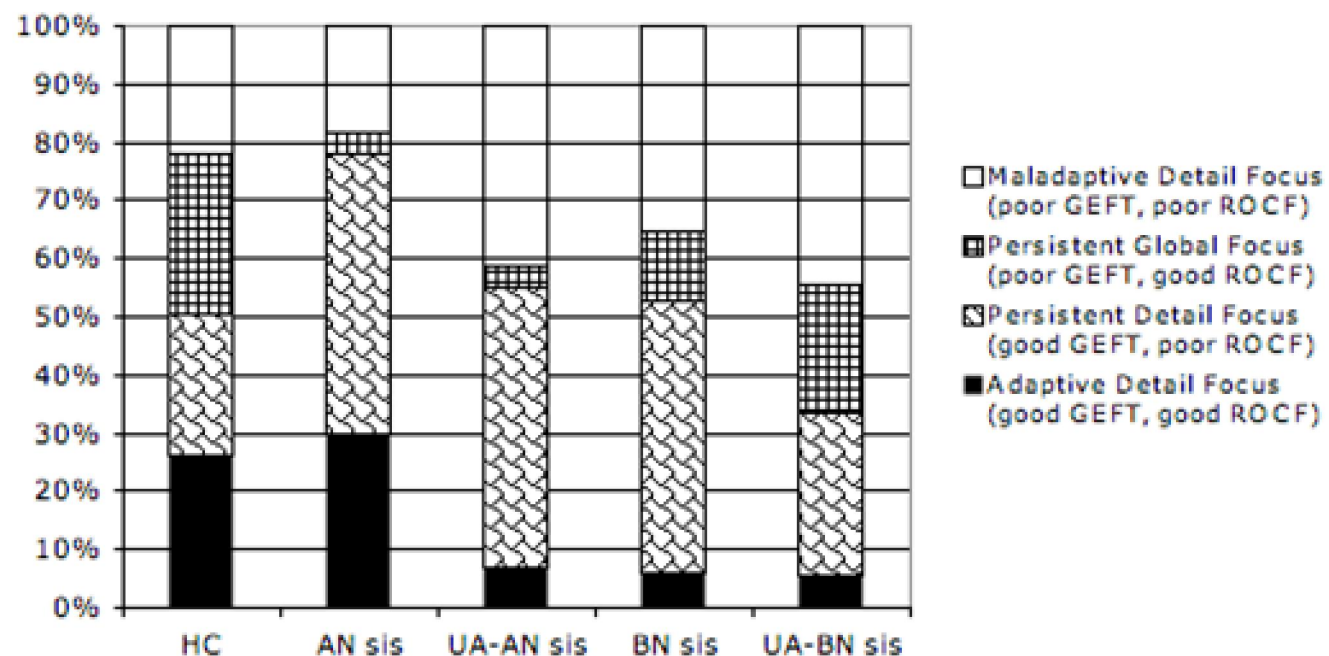
GEFT Group Embedded Figure Test; ROCF Rey-Osterrieth Complex Figure; HC Healthy Control; AN sis Anorexia Nervosa sister; UA-AN sis Unaffected Anorexia Nervosa sister; BN sis Bulimia Nervosa sister; UA-BN sis Unaffected Bulimia Nervosa sister

9.3.1   Set-shifting discussion

9.3.1.1   Analysis 1: ED discordant sister pairs

This initial analysis investigated set-shifting in sister pairs with and without an eating disorder (ED), to determine whether poor set-shifting found in study 1 in those with current ED was also present in family members unaffected by an ED. In this transdiagnostic approach including both AN and BN discordant sister pairs, no clear pattern across tasks emerged. The hypothesis that unaffected sisters would be worse on set-shifting tasks than HC was confirmed only for the WCST. In most cases, no significant differences were observed between sister pairs (in line with the first part of the hypothesis) or between unaffected sisters and HC (where the second part of the hypothesis predicted a significant difference).

Overall, three significant findings emerged: Unaffected sisters 1) made more perseverative errors on the WCST compared to HC, 2) completed significantly more categories than their ED sisters, and 3) reported significantly fewer perseverative illusions on the Haptic task compared to their ED sisters, while their results were comparable to those of HC women.

Given the finding of significantly more perseverative errors on the WCST in the UA-ED sister group compared to HC, it would have been expected that the UA-ED sister group would also complete fewer categories than the HC group. This is because more errors on this task delays its completion, which decreases the likelihood of all six categories being completed within the 128 card allocation. Such a pattern was observed amongst ED groups in study 1, where women with ED had a higher number of perseverative errors and fewer completed categories compared to HC women. However despite making a similar number of perseverative errors as their ED sisters, UA-ED sisters completed significantly more categories than their ED sisters. This suggests that UA-ED sisters made less random errors than their ED sisters, allowing them sufficient cards to complete all six categories despite their high number of perseverative errors. This may suggest lower levels of distractibility in healthy compared to ED sisters. Categories completed is a less specific measure of set-shifting ability compared to number of perseverative errors, as failure to complete all six categories of the WCST can be as much influenced by random errors or other extraneous task effects as by perseverative errors. Additionally, the inter quartile range for categories completed is at ceiling for both HC and UA-ED sisters, suggesting that it is not a good discriminator of neurocognitive performance.

Therefore the more sensitive set-shifting outcome on the WCST is that of perseverative errors. Taking these considerations into account, findings from the WCST suggest that, like their ED sisters, healthy sisters show a high level of perseverative responding indicative of poor cognitive flexibility when compared to control women. This provides evidence in support of set-shifting (as measured by the WCST) fulfilling criterion 4 of an endophenotype.

The WCST has often been used as an endophenotypic measure in the schizophrenia literature, where it has often been administered to non-psychotic siblings of those with schizophrenia. It should be noted that recent criticisms have suggested that the WCST is insensitive to heritability (Kremen, Eisen, Tsuang, & Lyons, 2007). However a meta-analysis of neurocognitive functioning in relatives of patients with schizophrenia reveal performance on the WCST to be significantly impaired in relatives compared to HC across 19 studies, with a small effect size of 0.29 (Sitskoorn et al., 2004). The current results suggest the WCST is a promising measure in the ED population with which to explore cognitive endophenotypes.

While no other significant findings were found between UA-ED sisters and HC women, a trend was seen on the TMT where both raw and balanced times for UA-ED sisters fell between those of HC and ED sisters. Additionally, of all three groups UA-ED sisters showed the highest number of errors on the Brixton task (indicating highest level of rigidity) although compared to the current HC group this effect size was small. On the Haptic task, UA-ED sisters were comparable in perceptual rigidity to the HC group, reporting significantly less illusions than their ED sisters. This finding differs to those on the WCST and the trends observed on the TMT and Brixton tasks, as it indicates that set-shifting in unaffected sisters is comparable to HC women and significantly better than ED sisters. However in previous chapters it has been suggested that the Haptic task, which has a strong perceptual component, may be tapping a construct separate to that of the other tasks (see 6.3.1.2). This hypothesis will be formally assessed in the following chapter.

One explanation for the lack of difference found between UA-ED and HC groups is that poor set-shifting may manifest differently in AN compared to BN families. Despite the larger statistical power obtained by combining AN and BN sister pairs, parsing out these two clinical groups may be more informative in examining the potential subtle differences between healthy relatives of those with the different subtypes of an ED and the general population.

9.3.1.2 Analysis 2: AN discordant sister pairs

The findings from the AN discordant sister pairs analysis very much follow those of the ED discordant sister pair analysis discussed above. This is perhaps not surprising, given that the majority of the ED discordant sample were AN sister pairs, and nearly half of those with lifetime BN had a history of AN. Two of the same three significant findings emerged, with unaffected AN sisters showing significantly more perseverative errors compared to HC, and completing significantly more categories than their AN sisters. A trend toward fewer illusions on the Haptic task compared to their AN sisters failed to reach significance (p=0.05).

Findings for the Brixton and Haptic tasks, which discriminated well between clinical and control groups in Study 1, did not directly confirm the current hypothesis. A small effect size was found between unaffected and AN sisters, rather than the hypothesised difference between unaffected sisters and HC. The direction of the finding indicates that unaffected sister performance is in line with HC performance, in that while AN sisters made significantly more errors on this task (Study 1), no impairment was seen for their unaffected sisters. This finding for the Haptic task is partially consistent with previous work (Holliday et al., 2006), where unaffected AN sisters displayed significantly fewer illusions than their AN sisters. However while in the Holliday sample a marked difference was found between healthy sisters and HC (ES=0.74), in the current sample not only was this difference smaller (ES=0.16) but the direction was reversed in that unaffected sisters displayed even fewer illusions then HC. Interestingly, should the current unaffected AN sister sample be compared with Holliday's HC sample, a direct replication would be seen in terms of the difference between healthy sister and HC performance (ES=0.62). This is due to the substantially smaller number of illusions reported by the HC group in the Holliday sample compared to the HC group in the current sample. As discussed in study 1 (see 6.3.1.2), this calls to question the performance of the current HC sample on the Haptic task; not only compared to the HC group of the Holliday study but also considering the consistently lower number of HC illusions (almost half of those found in the current study) observed in other studies previously employing this task (Tchanturia et al., 2001; Tchanturia et al., 2002; Tchanturia et al., 2004a).

Consistent with the Holliday sample, no differences emerged for the TMT. While the Brixton task also failed to reach significance, the direction of descriptive

statistics for this task highlights the need for further discussion. Study 1 indicated that the Brixton is a sensitive task. As hypothesised, no difference was found between AN and unaffected sisters in terms of errors made, indicating comparable performance on the task. However the hypothesised difference between unaffected sisters and HC was not present, despite the difference between AN and HC groups observed in Study 1. Further investigation of descriptive statistics shows that the number of errors made by unaffected sisters is medial to the errors made by AN sisters and HC, indicating that the anticipated directional relationship is present for the Brixton task (as with WCST perseverative errors), albeit that unaffected sister/HC analysis falls short of significance. Therefore it could be said that a trend in the direction of the hypothesis is seen for the Brixton task, in that affected and unaffected sisters showed a trend toward more errors (indicating more impaired set-shifting) compared to the HC group. In light of the endophenotype concept, these results provide a more consistent picture than those of Holliday et al. (2006), where similarly no significant differences were found between discordant sister pairs and the HC group on the Brixton task however the trend in mean scores differed substantially (most errors made by HC group, least errors made by unaffected sisters). As the distribution of mean scores across groups both in the current study and the Holliday paper is so tight, combined analysis of these samples may serve to more clearly illustrate group effects for the Brixton task. Overall, evidence for poor set-shifting ability in unaffected sisters of those with AN is provided by the WCST, with a trend toward support from the two other measures of cognitive rigidity.

9.3.1.3 Analysis 3: BN discordant sister pairs

Overall, like the AN discordant sister pairs some evidence for poor set-shifting is found, although the distribution of results across tasks differed. All tasks confirmed the first part of the hypothesis for BN, in that there was a negligible difference between sisters across all measures of set-shifting, indicating comparable performance (or impairment). No significant findings emerged on the WCST between unaffected BN sisters and HC, with a negligible effect size for perseverative errors. However, some evidence supporting the hypothesised difference between UA-BN sisters and HC was seen: In line with the results found in the BN split of Study 1 (see 6.2.4.4 and 6.2.6.4), significant findings emerged between UA-BN sisters and HC across TMT and Brixton tasks. In both cases, unaffected BN sisters showed more cognitive rigidity in their task performance than the HC group

(moderate effect sizes). This finding is particularly interesting for the Brixton task: Descriptive statistics were slightly lower (less errors) in the current subset of BN sisters than in the HC group. Despite this, UA-BN sisters made significantly more errors compared to HC using independent analysis (moderate effect size), but paired analysis between sisters revealed no difference between the two (negligible effect size) even with the wider distribution between group mean scores across sister groups. Such a finding indicates that sister pairs closely approximated each other in their performance on the Brixton task, in that both sisters performed well (e.g. 6/8 errors) or performed more poorly (e.g. 12/14 errors) on a consistent basis. Such an effect is masked when simply observing group mean scores, and illustrates the importance of paired analyses in this situation.

Like the Brixton task, UA-BN sisters showed more cognitive rigidity on the TMT task across both raw and balanced (B-A) variables with moderate effect sizes, with this difference reaching significance on the balanced variable only. Giving that this finding is evident in unaffected BN but not AN sisters, it is in line with the findings of study 1 where women with BN but not AN showed poorer set-shifting on the TMT. However in line with the findings of AUA-N sisters, UA-BN sisters also showed superior performance on the Haptic task compared to their BN sisters with the number of illusions paralleling those of the HC group. This again suggests that the Haptic task may be influenced by factors such as illness state, or an underlying impairment in basic perceptual processing in those with ED as illustrated by work from Grunwald and colleagues (Grunwald et al., 2001a; Grunwald et al., 2001b). Overall, evidence for poor set-shifting ability in unaffected sisters of those with BN is provided by two of the three cognitive measures of set-shifting. This is the first study to examine set-shifting in unaffected relatives of those with BN, therefore comparisons with prior findings cannot be made. Replication is required.

9.3.1.4 Analysis 4: Extreme scores

Extreme score analysis collapsed across tasks showed that unaffected BN sisters had a similar proportion of impaired shifting cases as their BN sisters, at approximately one third. In comparison, unaffected AN sisters showed few cases of impaired shifting, proportionate to that found in HC women and significantly less than that found in their AN sisters. This finding indicates that the trait of impaired shifting is more common in unaffected BN sisters, therefore providing stronger evidence for set-shifting as an endophenotype of BN than AN.

Further exploration of the criteria for impaired shifting may shed some light on the discrepancy between the unaffected sisters of those with AN and BN. Results presented in analysis 2 of this chapter show that UA-AN sisters differed significantly from HC on the WCST only. A moderate effect size was found on the WCST, while negligible to small group differences were seen on all other set-shifting variables. Comparatively, results presented in analysis 3 of this chapter show that UA-BN sisters differed significantly from HC on both the TMT and Brixton tasks, with negligible to small differences seen on the other two tasks. In order to be categorised as having impaired shifting, an extreme score (based on norms from the HC group) was required on at least two of the four set-shifting tasks. As UA-AN sisters differed moderately on one task, but UA-BN sisters differed moderately on two, it is perhaps not surprising that such a small proportion of UA-AN sisters were identified as meeting criteria for impaired shifting.

Unfortunately, when split across those with intact and impaired shifting, the number of unaffected ED sisters in the impaired shifting group was insufficient to conduct further analyses in terms of the relationship between set-shifting ability and both self-report clinical features and comorbidity. Despite this lack of power, it is interesting to note the lack of depression and multiple anxiety diagnoses in the impaired shifting group (n=8), while over a third of unaffected sisters with intact shifting had these diagnoses. This pattern is the opposite of that reported in study 1 (see 6.2.8.1), where those with current ED and impaired shifting showed a higher number of depression diagnoses and a significantly higher proportion of multiple anxiety diagnoses. No assessment of psychiatric illness in unaffected sisters of women with AN could be found in the literatures with which to compare this finding.

9.3.2   Coherence discussion

9.3.2.1   Analysis 1: ED discordant sister pairs

This transdiagnostic assessment aimed to investigate whether evidence for weak coherence was present in unaffected sisters of women with ED. It is the first study to explore the weak coherence hypothesis in 1st degree relatives of those with ED. Negligible to small group differences were found for the GEFT between both unaffected sisters and HC, and unaffected and ED sisters, with no comparisons reaching significance. On the ROCF, a significant difference was found on the coherence index where unaffected sisters of those with ED scored significantly lower

than HC (very large effect sizes), but no different to their ED sisters. ROCF copy accuracy scores did not differ between unaffected ED sisters and HC. Thus the findings across both the GEFT and ROCF differ in their direction, in that superiority of detail on one task (ROCF) is not present in the other (GEFT). Study 1 found notably different results on measures of coherence between AN and BN groups. Therefore the divergent results found in this transdiagnostic assessment may be due to the combination of unaffected sisters from two clinical groups, for whom the concept of weak coherence manifests differently. Exploration of the performance of the unaffected sisters of these two clinical groups is therefore most appropriately discussed separately, in the following sections.

9.3.2.2   *Analysis 2: AN discordant sister pairs*

The hypothesis for superior detail focus in the unaffected sisters of those with AN was confirmed across both measures of coherence. Unaffected sisters of women with AN were both faster than HC on the GEFT (small/moderate effect size), and displayed a significantly more detail focussed drawing style compared to HC on the ROCF (very large effect size). No differences were found between sisters. Copy accuracy scores on the ROCF did not differ between unaffected AN sisters and HC women. This indicates that global processing was intact in the AN sisters, as despite employing a different organisational strategy to HC women, unaffected sisters of those with AN were still equally accurate in the overall representation of the figure. Thus the low coherence index in healthy sisters of those with AN is indicative of a bias toward detail but intact global processing. Therefore results from both the GEFT and ROCF are suggestive of enhanced detail focus, confirming the hypothesis of superior detail focus in healthy sisters of those with AN. This pattern of findings is identical to that reported in Study 1 for women with current AN (see 6.2.4.4), providing evidence for detail focus as an endophenotype of AN.

It is interesting to note that while reaction time scores on the GEFT followed the expected trend (AN fastest, HC slowest with unaffected sisters falling between the two), on the ROCF healthy sisters had an even lower coherence index than AN sisters. This indicates that the drawing style of healthy sisters is even more fragmented and detailed than that of their affected sisters. While this trend did not reach statistical significance, the difference between sister groups approached a moderate effect size (0.39). Lower levels of perfectionism in unaffected compared to AN sisters may have contributed to the lower coherence index, in that less

perfectionism may mean less care is taken in the organisation of the drawing. However were this the case, lower copy accuracy may also be expected in the unaffected AN sisters. Subsequent correlational analysis between ROCF coherence index and self-report perfectionism in unaffected AN sisters did not reach significance (Spearman's $r(30)=0.36$, $p=0.06$). However the positive direction of the correlation indicates that unaffected sisters with lower perfectionism levels may be more likely have a lower coherence index or more heightened detail focus. Increasing statistical power will help to further explore this relationship.

This is the first study to investigate weak coherence in unaffected family members of those with AN. One such study investigating weak coherence in siblings of those with autistic spectrum disorder compared to control siblings is published in the autism literature, however methodological discrepancies make comparisons difficult. Happe et al. (2001) assessed mothers, fathers and siblings of autistic, dyslexic, and typically developing children with the block design task, EFT, visual illusion and sentence completion task. Family members of those with autism showed a trend toward weak coherence, with a significant group difference found for fathers but not siblings of those with autism compared to controls (Happe et al., 2001). As siblings have a closer genotype than parent/child dyads, it is surprising to see weak coherence represented more clearly in the parents of those with ASD than in their siblings. The small sample size in this study (13 siblings per group) likely impacted on findings and therefore replication is required. While a number of additional studies in the autism literature have investigated the neurocognitive profile of 1st degree relatives (e.g. Fombonne, Bolton, Prior, Jordan, & Rutter, 1997), these studies are limited by psychiatric rather than control comparison groups and, as above, small sample sizes. The results presented here for discordant AN sisters perhaps provides more compelling evidence for weak coherence as an endophenotype of AN than is yet seen in the autism literature.

9.3.2.3 Analysis 3: BN discordant sister pairs

In contrast to the hypothesis for healthy AN sisters, it was hypothesised that healthy BN sisters would show no superiority with detail but poor global processing across measures of weak coherence. Findings from the GEFT confirmed that a bias toward detail was not present in UA-BN sisters, with no difference found between UA-BN sisters and HC (negligible/small effect size). UA-BN sisters were on average one second slower on this task than HC. No differences were found between healthy

and BN sister pairs. On the ROCF, the same pattern as that observed amongst UA-AN sisters was found where UA-BN sisters had a significantly lower coherence index for their drawing compared to HC, which did not differ compared to their affected sister. However unlike the AN group, healthy BN sisters scored significantly lower on copy accuracy compared to HC. As discussed in study 1 (see 6.3.2.2), this thesis proposes that coherence index is meaningful only in the context of copy accuracy scores. Where accuracy scores are high (i.e. comparable with HC) this suggests intact global integration. Therefore, adoption of a fragmented drawing style (as reflected by a low coherence index) in the context of high accuracy is indicative of a bias toward detail. In contrast, significantly lower copy accuracy score indicate poor global processing, in the context of which a fragmented drawing style (low coherence index) is a factor of this poor global integration rather than a bias toward detail. Thus the hypothesis of poor global integration in healthy sisters of those with BN has gained support across tasks here, implicating poor global integration as an endophenotypes of BN.

9.3.2.4 *Analysis 4: Extreme scores*

The performance of affected and unaffected sister pairs was collapsed across coherence tasks according to task strategy. The most highly endorsed quadrant for unaffected AN sisters was persistent detail focus, where regardless of task demands nearly half of all UA-AN sisters employed a detail focussed processing style. This finding is consistent with that of the AN sisters presented here and those with current AN reported in study 1 (see 6.2.8.2), who also largely endorsed the persistent detail focus quadrant. This consistent findings across populations provides strong evidence for a natural bias toward detailed or local processing as an endophenotype of AN. While unaffected BN sisters also showed a marked number of cases with persistent detail focus, this quadrant was not the most highly endorsed at just over 25%. Rather, nearly half of UA-BN sisters showed what was termed maladaptive detail focus, in that a fragmented style was employed when it was not advantageous (ROCF) but local processing was not employed when it would have assisted task performance (GEFT). This pattern however is not surprising, given that their BN sisters showed nearly twice the proportion of cases with this style compared to AN sisters, and women with current BN reported in study 1 also showed high endorsement of this category (see 6.2.8.2). As discussed in 6.3.2.4, it is likely that these cases with a slow GEFT time but a low coherence index on the ROCF are those with poor planning

and impulsivity, as evidenced by a low ROCF accuracy score. In such cases, these results are likely more indicative of poor global integration, as evidenced by a fragmented and inaccurate picture completion, and normal levels of detail as evidenced by taking a similar amount of time to find hidden shapes as HC. It is therefore fitting that high numbers of this quadrant are observed in the BN spectrum, extending to unaffected 1st degree relatives.

As mentioned above in 9.3.1.4, there were insufficient numbers of unaffected BN sisters with adaptive detail focus (n=3) for a comparison of adaptive/persistent detail focus across self-report and psychiatric diagnoses to be made.

9.4   General Conclusions

Study 4 has shown that across both AN and BN discordant pairs, unaffected and ED sisters do not differ notably from each other on measures of set-shifting and weak coherence. Study 1 provided evidence of those with both AN and BN exhibiting difficulties with shifting set. While the transdiagnostic assessment of discordant ED sister pairs revealed little substantial evidence, when split the unaffected sisters of both AN and BN women showed some evidence of poor set-shifting across different neurocognitive tasks. The expected trend of poor flexibility was seen across most tasks, however findings reached significance on the WCST only for unaffected AN sisters, and the TMT and Brixton tasks for unaffected BN sisters. Thus the current study provides some evidence in favor of the hypothesis that unaffected sisters, like their ED sisters, show difficulties with cognitive flexibility. The Haptic task does not show strong evidence as an endophenotypic measure. When collapsed across measures of flexibility, there was limited evidence for poor set-shifting in UA-AN sisters but moderate evidence for UA-BN sisters.

The current study also provided evidence in favor of the hypothesis that unaffected sisters, like their ED sisters, exhibit weak coherence when processing information. However the manifestation of weak coherence differed in line with the ED sister's diagnosis. Consistent with the results for women with AN in study 1, unaffected sisters of those with AN showed a bias toward local processing. Consistent with the results for women with BN in study 1, unaffected sisters of those with BN showed poor global integration and no evidence for a bias toward local processing. When collapsed across both tasks, a consistent pattern emerged where nearly half of UA-AN sisters fell into the category of persistent detail focus, while

nearly half of the UA-BN sisters showed evidence of poor global integration. These consistent findings across studies and methodologies lends support to the conceptualisation of different aspects of the weak coherence hypothesis being implicated as endophenotypes for AN and BN.

Overall, study 4 provides strong evidence for weak coherence as an endophenotype of ED (particularly for AN) and moderate evidence for poor set-shifting ability as an endophenotype of ED (particularly for BN).

9.4.1   *Limitations*

Three main limitations are of note regarding the current study. Firstly, nearly half (40%) of sisters in the BN sample had a history of AN. Therefore, findings from the BN group may have been clouded by the AN history of the group. This thesis grouped participants based on lifetime diagnosis as outlined by the DMS-IV, in that the presence of BN 'trumped' the presence of AN. However an equally valid method of classification would group those with a lifetime underweight ED (AN and mixed AN/BN) and those with lifetime normal weight ED (BN only). Such a classification was explored in Study 1, however the number of sister pairs that would have formed the pure BN group (n=12) meant that this was outside the scope of the current study. This approach may be more appropriate when assessing endophenotypes, as lifetime phenotypic classification may more appropriately tap underlying biological traits compared to classification based on a clinically determined hierarchy of behaviours. Further work in this area should target families where BN but not AN is present, in order to more accurately investigate the manifestation of weak coherence and poor set-shifting as candidate endophenotypes of BN.

Secondly, due to recruitment difficulties the current study was able to collect a sample size of just 20 sister pairs discordant for BN compared to the 30 sister pairs discordant for AN. Power analysis presented in chapter 5.6 indicated that approximately 30 participants per group were required in order for statistical power to detect differences. It is therefore possible that the findings presented here are underpowered, and may have differed should a larger sample have been collected. It is unlikely that such an increase in power would dramatically change weak coherence results in healthy BN sisters, given the direction of descriptive statistics.

Finally, unaffected sisters of those with ED were deemed 'unaffected' based on the lack of a lifetime ED diagnosis (be it AN, BN or EDNOS). However

unaffected sisters were not screened based on other psychiatric diagnoses, which resulted in a number of sisters with current depression and anxiety diagnoses. It is well recognised that set-shifting is a candidate endophenotype not restricted to the ED, but also found in other psychiatric populations such as those with mood disorders (Bora et al., 2008). Therefore, the presence of poor flexibility in sisters with a lifetime psychiatric diagnosis may be explained by the presence of their own illness, rather than that of their ED sister. Such a manifestation may be a factor of shared vulnerability to the candidate endophenotype and therefore psychiatric illness in general. Exploration of cognitive flexibility in two subsets of unaffected sisters-those with a psychiatric history and those with no psychiatric history- would serve to further inform this question.

.

10   The combined effect: Exploring the relationship between poor set-shifting and weak coherence

10.1   Background

This thesis has explored poor set-shifting and weak coherence in those with current and past eating disorders (ED), in addition to unaffected sisters of those with an ED. Study 1 found that extreme cases of poor set-shifting in women with a current ED (as determined by healthy control [HC] data) were associated with poor prognostic factors such as longer duration of illness, more severe eating behaviours, higher self-harm, lower self-esteem and more comorbid anxiety diagnoses. Similarly, poor prognostic factors were found in those with a current ED and persistent detail focus, for example lower self-report depression and higher rates of social phobia, specific phobia, and number of comorbid anxiety diagnoses.

As yet, the relationship between these two aspects of neurocognitive functioning has not been directly assessed. In the discussion of the extreme scores analysis in Study 1, a hypothesis was posed regarding a potential interaction between the traits. It was suggested that, because of the ability to change strategy across measures of coherence according to task demands, participants with adaptive detail focus may also be those with intact cognitive flexibility (see 6.3.2.4). The aim of this chapter is to investigate whether any direct relationship exists between these two information processing styles, and whether having both traits rather than one has a compounding effect on eating symptoms, their severity and prognostic factors.

10.1.1 Hypotheses

It is hypothesised that 1) a relationship will exist between set-shifting and coherence ability, specifically those with intact shifting will show higher levels of adaptive detail focus compared to those with impaired shifting, and 2) those presenting with both traits compared to one trait will show more severe and persistent clinical features and comorbidity.

10.2   Method

10.2.1 Participants

Participants were 270 women with and without an ED previously included in chapters 6-9. The clinical composition was as follows; 68 anorexia nervosa (AN), 30 bulimia nervosa (BN), 30 recovered AN, 30 unaffected sisters of women with AN

(UA-AN), 20 unaffected sisters of women with BN (UA-BN). Some analyses use a subset of the data, for example women with lifetime ED only (n=128).

10.2.2 Statistical methods

To investigate the direct relationship between tasks, spearman's rank correlations were run between set-shifting and coherence variables. Main outcomes for each task (TMT B-A; WCST perseverative errors; Brixton errors; Haptic perseverations; GEFT median time; ROCF coherence index) were then subject to a principal components analysis (PCA) using varimax rotation, extracting factors with eigenvalues greater than 1. Missing values were excluded casewise. Factor loadings less than 0.3 were suppressed.

Those with extreme scores on set-shifting and coherence neuropsychological tasks were identified using the procedure outlined previously (see 6.2.7.2). Chi-square tests were used to analyse differences in the frequency of women with lifetime ED and impaired shifting across each of the four

All current ED cases were split into those with both impaired shifting and persistent detail focus, and those with one trait or the other. A chi-square test was employed to investigate differences in the frequency of those presenting with both traits by diagnostic group. Groups were then compared across demographic, clinical, and self-report clinical features using independent samples t-tests (Mann-Whitney U tests when data was not parametric) and comorbidity using chi-square tests.

10.3   Results

10.3.1 Relationship between set-shifting and coherence

10.3.1.1 Correlational analysis

Correlations were calculated between each of the main outcomes of set-shifting and coherence tasks for those with a lifetime ED diagnosis. GEFT median time correlated significantly with WCST perseverative errors ($r(110)=0.24$, $p=0.01$) and Brixton errors ($r(118)=0.20$, $p=0.03$), indicating that those with faster times on the GEFT made less errors on set-shifting tasks. ROCF coherence index showed a significant negative correlation with both Brixton errors ($r(126)=-0.19$, $p=0.04$) and Haptic illusions ($p(123)=-0.28$, $p<0.01$), indicating that those with a lower coherence index made more errors/illusions.

10.3.1.2 Principal components analysis (PCA)

PCA with extraction based on eigenvalues greater than 1 revealed 2 components, with cross-loadings for the TMT only. The 2-factor solution explained 49.67% of the total variance in scores. See Table 58 for the rotated component matrix. The Brixton, WCST and TMT clustered together with the GEFT on factor 1, while the ROCF clustered with the TMT and inversely with the Haptic task on factor 2. Both factors showed poor internal reliability (factor 1 α=0.49; factor 2 α=0.09), which was not improved by eliminating any of the items.

10.3.1.3 Analysis using extreme scores

Initial frequency analysis explored the prevalence of impaired shifting across the four quadrants of coherence (adaptive detail focus, persistent detail focus, persistent global focus, maladaptive detail focus) in women with lifetime ED (see Figure 30). The group with adaptive detail focus had the fewest cases of impaired shifting (10.3%). This was well below the average proportion of lifetime ED impaired shifting cases (30.2%). The highest rate of impaired shifting was seen in those with maladaptive detail focus (53.8%). Those in the persistent detail and persistent global focus quadrants showed similar rates (34%; 40% respectively).

The rate of those with impaired shifting was significantly lower in the adaptive detail focus group compared to those with persistent detail focus (χ^2(1)= 5.10, p=0.02; Cohen's d=0.53), persistent global focus (χ^2(1)= 4.21, p=0.04; Cohen's d=0.71), and maladaptive detail focus (χ^2(1)= 11.63, p<0.01; Cohen's d=1.05).

10.3.2 Simultaneous presentation of impaired shifting and persistent detail focus

10.3.2.1 Incidence across diagnoses

Analysis of frequency data for current ED with impaired shifting, persistent detail focus, and both traits simultaneously is presented in Figure 31. The BN group showed the highest number of cases with both traits at 22%, followed by ANBP at 18% and ANR at 9%. Most of those recovered from AN who showed impaired set-shifting (10%) also displayed persistent detail focus (8%). The ANR and recovered AN groups showed moderately less cases compared to ANBP (Cohen's d=0.43; 0.42) and BN (0.47; 0.47). There was no significant difference in the frequency of those with both traits across ED diagnoses (χ^2(3)=5.65, p=0.13). One HC participant met criteria for both traits. No unaffected sisters of women with AN or BN met criteria for both traits.

Table 58: Rotated component matrix for principal components analysis

	Component	
	1	2
Brixton errors	0.73	
GEFT median time	0.68	
WCST perseverative errors	0.64	
TMT balanced variable (3-2)	0.57	0.30
ROCF coherence index		0.83
Haptic illusions		-0.67

10.3.3 Relationship with demographic and clinical features

Groups with one and both traits were demographically similar and showed negligible differences on most clinical features (see Table 59). The group with both traits had a significantly higher score with a large effect size on YBC-EDS rituals. This indicates more severe eating behaviours measured at the worst phase of the illness (e.g. ritualised eating, food preparation, body checking and exercising). In line with this finding, moderate differences were also seen on YBC-EDS preoccupations, and lowest ever BMI, again indicating a more severe illness. Comparisons on self-report measures showed no significant group differences (see Table 60). One moderate difference was found, where those with both traits had a higher score on the concern over mistakes subscale of the Frost multidimensional perfectionism scale. No other effect sizes were greater than small.

10.3.3.1 Relationship with comorbidity

No significant group differences on comorbidity were seen (see Table 61). Those with both traits had moderately higher levels of obsessive-compulsive disorder (OCD), generalised anxiety disorder (GAD) and self-harming behaviours. Women with both traits were 2.56 times more likely to have a depression diagnosis, although the effect size was small.

10.4 Discussion

This chapter investigated for the first time the relationship between set-shifting and coherence in women with lifetime ED. Those with adaptive detail focus as their coherence strategy show the lowest rates of impaired set-shifting, confirming the hypothesis raised in study 1. This finding shows that nearly 90% of women with a lifetime ED and adaptive detail focus also show a flexible cognitive profile. This flexibility enables them switch from a local processing style to a more global style as the task or situation demands. In contrast, those with persistent detail focus or persistent global focus are more biased in their approach to a given task, in that they employ the same strategy regardless of the task or situation demands. This is evidenced by the higher rate of those with impaired flexibility (approximately one third) in women with these coherence strategies. The highest rates of impaired shifting were seen in the maladaptive detail focus group, where over 50% of cases showed marked cognitive rigidity.

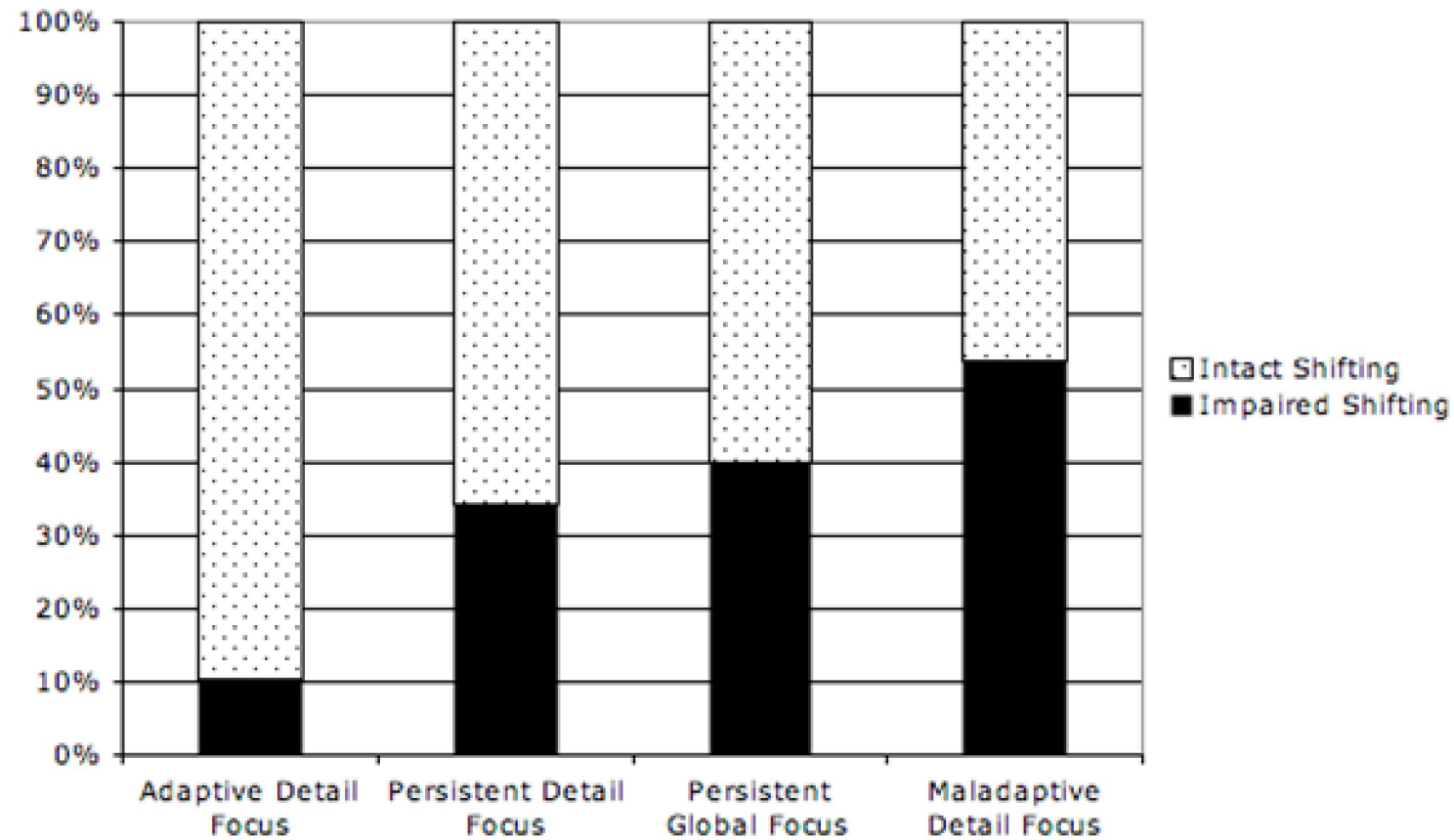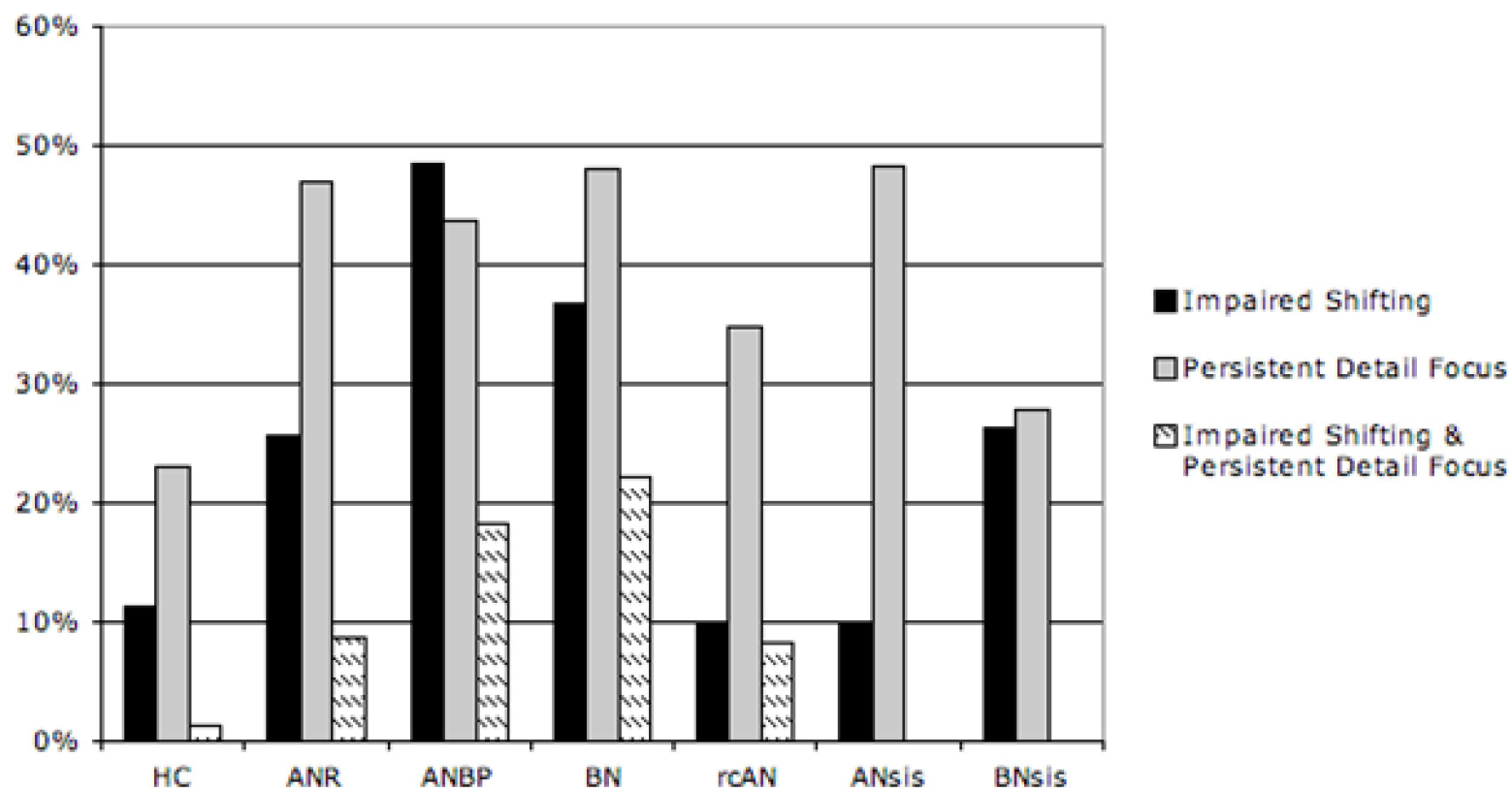Figure 30: Set-shifting ability by coherence strategy in lifetime ED

Figure 31: Frequency of impaired shifting and persistent detail focus both independently and combined by diagnostic group

HC Healthy Control; ANR Restricting type Anorexia Nervosa; ANBP Binge/purging type Anorexia Nervosa; BN Bulimia Nervosa; rcAN Recovered Anorexia Nervosa; ANsis Unaffected Anorexia Nervosa sister; BNsis Unaffected Bulimia Nervosa sister

Table 59: Demographic and clinical features for current ED with both or one of impaired shifting and persistent detail focus

	One trait	Both traits	Test statistics			
	(n=48)	(n=15)	t	MW	p	Cohen's d
Age	24.73 (7.69)	24.87 (5.00)	-0.07	-	0.95	0.02
Education (years)	15.39 (2.89)	15.70 (2.97)	-0.36	-	0.72	0.11
Current medication [1]	0 (0-1)	1 (0-1)	-	92.0	0.84	0.07
Current ED severity	2.56 (1.07)	2.33 (1.35)	0.68	-	0.50	-0.20
Age of onset	16.56 (3.45)	16.07 (3.10)	0.67	-	0.50	-0.15
Duration of illness [eq]	7.88 (6.96)	7.50 (4.25)	0.20	-	0.85	-0.06
YBC preoccupations [eq]	12.23 (2.33)	13.53 (2.17)	-1.91	-	0.06	0.57
YBC rituals	11.05 (3.49)	13.73 (2.28)	-2.78	-	<0.01	0.82*
	(n=36)	(n=9)				
Current BMI	18.43 (2.56)	17.51 (2.63)	-	-		0.36
Lowest BMI	15.36 (5.42)	12.47 (1.97)	-	-		0.58

MW Mann-Whitney U Test; ED Eating Disorder; YBC Yale-Brown-Cornell Eating Disorder Scale; BMI Body Mass Index
[eq] Equal variances not assumed (Levene's test for equality of variance <0.05)
[1] Current medication data not normally distributed, therefore descriptive statistics presented are median with interquartile range in parentheses
* Comparison significant at 0.05 level

Table 60: Self-report clinical features for women with current ED and both or one of impaired shifting and persistent detail focus

	One trait (n=46)	Both traits (n=15)	Test statistics t	MW	p	Cohen's d
HADS anxiety	12.29 (4.55)	13.00 (4.84)	-0.51	-	0.61	-0.15
HADS depression	6.77 (4.33)	8.20 (4.52)	-1.10	-	0.28	-0.32
OCI-R total [eq]	21.16 (16.06)	20.23 (11.52)	0.21	-	0.84	-0.07
Hording	3.56 (3.06)	3.73 (3.01)	-0.20	-	0.85	-0.06
Checking	2.22 (2.81)	2.67 (2.06)	-0.56	-	0.58	-0.17
Ordering	4.24 (3.95)	4.63 (3.06)	-0.35	-	0.73	-0.10
Neutralising	2.36 (3.08)	1.93 (1.98)	0.50	-	0.62	0.15
Washing	2.58 (3.46)	1.67 (2.19)	0.96	-	0.34	0.28
Obsessing	6.20 (4.18)	5.60 (4.14)	0.48	-	0.63	0.14
Rosenberg self-esteem	11.30 (5.39)	11.05 (5.35)	-0.94	-	0.35	0.05
Frost Perfectionism	95.62 (17.49)	100.30 (14.47)	-0.94	-	0.35	-0.28
Concern mistakes	31.85 (9.17)	35.63 (7.27)	-1.45	-	0.15	-0.43
Personal standards	26.65 (5.00)	27.03 (5.80)	-0.25	-	0.81	-0.07
Doubting	13.83 (3.31)	14.60 (3.11)	-0.79	-	0.43	-0.24
Organisation	23.28 (5.37)	23.03 (5.29)	0.16	-	0.88	0.05
ChiRP total [1]	6 (3-11)	5.5 (3.5-12.25)	-	209.5	0.77	0.08

Perfectionism [1]	2 (0-4)	3 (1-4)	-	221.5	0.41	0.23
Inflexibility [1]	2 (1-3)	2 (1.5-4)	-	252.0	0.87	0.05
Order/symmetry [1]	1 (0-4)	1 (0-5.25)	-	275.0	0.62	0.12

MW Mann-Whitney U Test; HADS Hospital Anxiety and Depression Scale; OCI-R Obsessive-Compulsive Inventory-Revised; CHiRP Childhood Retrospective Perfectionism Questionnaire

[eq] Equal variances not assumed (Levene's test for equality of variance <0.05)

[1] Data is not normally distributed, therefore median is presented with upper and lower quartiles in parentheses. All other descriptive statistics mean with standard deviation in parentheses

Table 61: Lifetime psychiatric diagnoses for current ED with both or one of impaired shifting and persistent detail focus

	One trait (n=44)	Both traits (n=15)	χ^2	p	Odd's ratio	Cohen's d
Anxiety Disorders						
OCD	45.5%	66.7%	2.01	0.16	2.18	0.38
OCPD	29.5%	26.7%	0.05	0.83	0.87	-0.06
Panic Disorder	37.0%	26.7%	0.53	0.47	0.62	-0.19
Social Phobia	45.7%	46.7%	0.01	0.95	1.04	0.03
Specific Phobia	30.4%	33.3%	0.04	0.83	1.14	0.05
PTSD	13.0%	6.7%	0.45	0.50	0.48	0.18
GAD	10.9%	26.7%	2.24	0.13	2.98	0.40
BDD	0%	0%	-			
Multiple Diagnoses [1]	50.0%	60.0%	0.45	0.50	1.5	0.18
Mood Disorders						
MDD	71.7%	86.7%	1.36	0.24	2.56	0.31
Bipolar Disorder	6.5%	0%	1.03	0.31	-	0.27
Dysthymia	2.2%	0%	0.33	0.57	-	0.15
Self-harm	46.7%	71.4%	2.63	0.11	2.86	0.43
Substance Disorders						

Alcohol Abuse	15.2%	20.0%	0.19	0.66	1.39	0.11
Alcohol Dependence	17.4%	13.3%	0.14	0.71	0.73	-0.10
Substance Abuse	8.7%	13.3%	0.27	0.60	1.62	0.14
Substance Dependence	15.2%	13.3%	0.03	0.86	0.86	-0.05

AOO Age of onset; DOI Duration of illness; OCD Obsessive-compulsive disorder; OCPD Obsessive-compulsive personality disorder; PTSD Post-traumatic stress disorder; GAD Generalised anxiety disorder; BDD Body dysmorphic disorder; MDD Major depressive disorder

[1] Split into those with no or one anxiety diagnosis and those with 2 or more (multiple) anxiety diagnoses

Further frequency analysis revealed that the simultaneous presentation of both impaired shifting and persistent detail focus was present in those with lifetime ED. Those with BN showed the highest percentage of cases with both traits, where moderate (but not significant) differences were found between the four clinical groups. Most cases with both traits fell on the BN spectrum, suggesting that those with both impaired shifting and persistent detail focus are less likely to have ANR. Those with ANR who do have both traits are perhaps more likely to transition to ANBP or BN. One spurious HC case met criteria for both traits. This participant reported high obsessive-compulsive behaviours but a healthy BMI and no indication of disordered eating. This case may illustrate that the relationship between OCD and the presence of both traits (see 10.3.3.1) is also relevant to the general population. As unaffected sisters of those with ED displayed some of the same traits as their ED sisters but never both impaired set-shifting and persistent detail focus together, it may be that the compounding effect of both neurocognitive traits contributes to the development of an ED.

When women with current ED were split into those with both traits and those with one, few group differences emerged. Those that did were consistent with the hypothesised poorer prognostic factors in the group with both traits. Higher YBC rituals (significant), preoccupations, and a moderately lower 'worst' BMI were reported, suggesting those with both impaired shifting and persistent detail focus had a more severe illness than those with one trait only. Those with both traits also showed moderately higher levels of lifetime GAD and self-harm, and were over two times more likely to have a lifetime depression or OCD. Negligible to small differences were seen across all other comparisons.

Significant correlations were seen between set-shifting and coherence variables for all but the TMT. Good performance on each coherence task (i.e. fast GEFT, global ROCF) was associated with a flexible cognitive style. This suggests an overall style of adaptive detail focus across diagnoses. This finding is inconsistent with the results of study 1, where adaptive detail focus was the least endorsed coherence strategy across diagnoses (see 6.2.8.2), however it is consistent with results from the recovered AN cohort (see 7.3.4.3). It is likely that the heterogeneous nature of the transdiagnostic lifetime ED group analysed here has influenced the results. While set-shifting is a linear concept in that one falls either more toward the intact or impaired end of the spectrum, coherence being a more dimensional concept

means that it's relationship with set-shifting is perhaps not best understood using a correlational method. This is because a participants' coherence ability is more accurately understood as one of four dimensions (as determined by both coherence tasks), rather than as a linear measurement of one task as is assumed when correlating two variables. A correlational method may be more appropriate were a more pure measure of global processing employed to compare with the GEFT, a measure of detail focus.

This same rationale is likely why the exploratory principal components analysis contributes little to the understanding of this data. Cognitive set-shifting tasks clustered together in factor 1 along with the GEFT, the more cognitive of the coherence tasks. Factor 2 negatively clustered the ROCF and Haptic tasks, the more creative of the neuropsychological measures where drawing and perceptual skills are required. While the two factors explained nearly half of the variance in scores, low cronbach's alphas within both factors suggests poor reliability. Therefore, at least for this initial investigation into the relationship between set-shifting and coherence, frequency analyses may provide a more cleaner means with which to explore this data.

10.4.1 Conclusions

In sum, this chapter has conducted the first investigation into the relationship between impaired shifting and detail focus. Exploration of this relationship using statistical techniques such as factor analysis is difficult, given the dimensional nature of coherence and the tasks employed in the current thesis. A small proportion of those with lifetime ED display both traits (15%), with those with BN showing the highest rates (22%). Where both traits were seen simultaneously, a more severe ED was noted in terms of lowest ever BMI and eating related preoccupations and rituals. Additionally, the likelihood of comorbid diagnoses such as OCD, GAD, depression and self-harm was more than doubled.

11   Summary and Conclusions

11.1   Revising the objective

The aim of this thesis was to further understanding of aspects of neurocognitive functioning that may serve as risk and maintaining factors of eating disorders (ED). Focus was placed on two aspects of cognitive style: set-shifting (the ability to be flexible in ones cognitive and behavioural processes, allowing for adaptation to changes in and/or novel situations), and weak coherence (the natural tendency to prioritise processing of local over global elements, which can result in difficulty seeing the bigger picture). This thesis had two main objectives with regard to these traits; 1) to explore whether set-shifting and weak coherence fulfil the criteria for endophenotypes of ED, and 2) to define the clinical validity of these traits both in isolation and when presenting simultaneously, in terms of prognosis and comorbidity.

In order to test the first objective, both traits were investigated as candidate endophenotypes of ED by addressing four of the five criteria that have been proposed as indicators of a psychiatric endophenotype (Gottesman & Gould, 2003). The criteria addressed were; 1) association of the trait with the illness (chapter 6), 2) state-independence of the trait (chapter 7), 3) presence of the trait in affected relative pairs (chapter 8), and 4) presence of the trait in unaffected relatives of those with the illness (chapter 9). Data for the final heritability criteria of an endophenotype was collected in the form of blood samples, however DNA analysis was outside the scope of this thesis.

In order to test the second objective, chapters 6-9 also explored the incidence of these traits in the ED population, in addition to their relationship with clinical features and comorbid psychiatric diagnoses. Finally, the relationship between the two neurocognitive traits was explored (chapter 10).

This final chapter will summarise and discuss the main findings obtained in this thesis, and set them in the context of the search for endophenotypes in ED and potential links with neurobiology. This is followed by strengths and weaknesses of the study with a focus on methodology, and the current study's limitations. To conclude, discussion of how the current findings can be applied to a variety of clinical settings is presented, and future research directions for the field are outlined.

11.2 Summary of neuropsychological findings

11.2.1 Set-shifting findings

11.2.1.1 Set-shifting as an endophenotypes of ED

Set-shifting in women with ED was investigated using a transdiagnostic approach across the first, third and last criteria outlined above (framed as hypotheses). The first hypothesis was that women with current ED would show poor set-shifting compared to HC women. This hypothesis was confirmed on the TMT, WCST, Brixton and Haptic tasks with small to moderate effects. This finding is consistent with previous studies of AN and BN as outlined in the meta-analysis in chapter 4. The third hypothesis was that sister pairs concordant for an ED would not differ in their set-shifting ability. This hypothesis was confirmed, in that sister pairs showed a consistent pattern across tasks in Study 3. No previous studies have investigated set-shifting or indeed neurocognition in sister pairs or family members concordant for an ED.

The last hypothesis was that unaffected sisters of those with ED would show poor set-shifting compared to HC women. This was confirmed to a level of significance on the WCST only, where unaffected sisters were more rigid with a moderate effect. However, small effects on the Brixton and TMT tasks showed a trend in the direction of rigidity for unaffected sisters. Effect sizes across these three tasks were equivalent to or greater than half the effect size seen between ED and HC women in Study 1. As siblings share approximately 50% of their genes, the presence of an endophenotype in an affected sister should logically be represented at half that level in the unaffected sister. Therefore evidence for the presence of poor set-shifting in unaffected sisters of women with ED is present at the expected rate of half the effect of ED sisters across the TMT, WCST and Brixton tasks.

11.2.1.2 Set-shifting as an endophenotype of AN

Set-shifting in AN was investigated across the first, second and last criteria outlined above. The first hypothesis was that women with current AN would show poor set-shifting compared to HC women. This hypothesis was confirmed, with results from Study 1 showing poor set-shifting in AN compared to HC women on the WCST, Brixton and Haptic tasks. This finding is comparable to AN studies outlined in chapter 4 with an outpatient sample with similar BMI (Holliday et al., 2005; Steinglass et al., 2006) and those with more severe inpatient samples (Fassino et al., 2002; Ohrmann et al., 2004; Tchanturia et al., 2004a). No effect of BMI was found

within the current AN sample, indicating that poor set-shifting was not a factor of malnutrition. This is the first time neuropsychological performance has been examined across the different subtypes of AN, therefore it was interesting to note that poor set-shifting was particularly evident in women with a mixture of AN and BN behaviours compared to restricting type AN.

The second hypothesis was that women recovered from AN would not differ in set-shifting ability compared to current AN, but would show poor set-shifting compared to HC women. This hypothesis was confirmed only in part. Evidence for persistent set-shifting difficulties was found in the recovered AN group compared to HC on the WCST only. However other tasks showed results in line with HC performance (negligible to small effects), suggesting a degree of improved cognitive flexibility with illness recovery. This finding of partial improvement is consistent with previous investigations of set-shifting in women recovered from AN (Pendleton-Jones et al., 1991; Tchanturia et al., 2004b), suggesting that cognitive rigidity may, at least in part, be a state effect.

The last hypothesis was that unaffected sisters of those with AN would show poor set-shifting compared to HC women. While statistically conclusive evidence was found on the WCST only, support for this hypothesis was provided across the other tasks in that unaffected sisters of those with AN fell consistently between the results of their AN sisters and HC. Negligible to small differences were seen between unaffected sisters and both AN sisters and HC, indicating that subtle differences were present between the groups. This finding is consistent with the previous study of set-shifting in unaffected first degree relatives, where unaffected sisters were significantly worse on the CatBat task, but not the TMT or Brixton tasks (Holliday et al., 2005). Subtle effects in unaffected relatives may be best examined using effect sizes rather than multivariate analyses.

Overall, findings provided moderate evidence across criteria for set-shifting as an endophenotype of AN.

11.2.1.3 Set-shifting as an endophenotype of BN

Set-shifting in BN was investigated across the first and last criteria of an endophenotype. The first hypothesis was that women with current BN would show poor set-shifting compared to HC women. This hypothesis was confirmed across the TMT, WCST and Haptic tasks. A consistent finding remained when groups were split by normal weight BN only and BN with a history of AN. This finding is

consistent with the few previous studies that have examined set-shifting in BN populations (Pendleton-Jones et al., 1991; Tchanturia et al., 2002; Tchanturia et al., 2004a), none of which have split by the presence of lifetime AN. This study provides the second replication of moderate effects on the TMT and Haptic tasks for BN, and is the first time results have been presented for BN on the WCST.

The last hypothesis was that unaffected sisters of those with BN would show poor set-shifting compared to HC women. This hypothesis was confirmed on some tasks, where unaffected sisters of those with BN showed significantly poorer set-shifting on the TMT and Brixton tasks compared to HC with moderate effects. Negligible to small differences were seen across the other tasks. Discordant sisters did not differ significantly from each other, with small to moderate effect sizes. This was the first study to explore the neuropsychological profile of unaffected sisters or relatives of those with BN.

Overall, findings provide moderate evidence for set-shifting as an endophenotype of BN. Investigation of this trait in women recovered from BN is still required.

11.2.1.4 Incidence and impact of set-shifting on illness factors

An investigation of cases with 'extreme scores' across set-shifting tasks (above 1 SD of HC mean) was conducted. Overall, 37% of those with a current ED had 'impaired shifting' (extreme scores on 2 or more set-shifting variables). Across diagnoses, women with ANBP were most likely to have 'impaired' shifting with nearly half of ANBP cases fulfilling this criterion. Women with ANR showed 26% with impaired shifting, while 37% of women with BN showed impaired shifting. All current ED groups had significantly more cases of impaired shifting compared to the HC group (11%). In both the recovered AN group and unaffected sisters of those with AN, only 10% of cases met criteria for impaired shifting. For unaffected BN sisters, this number was more in line with that of their BN sisters (30%). Figure 32 depicts the weighted proportion of impaired shifting cases across participant groups.

These findings have outlined that poor set-shifting is not universal, but rather is characteristic of a subgroup of individuals with ED. This observation was also made by Lauer and colleagues (Lauer et al., 2002) in a prospective study, however they found no relationship between neurocognition and clinical features or subsequent recovery. Unlike the null findings of Lauer et al. (perhaps due in part to their lack of power; n=26), this thesis found that impaired cognitive flexibility was

Figure 32: Weighted proportion of impaired shifting cases across participants groups
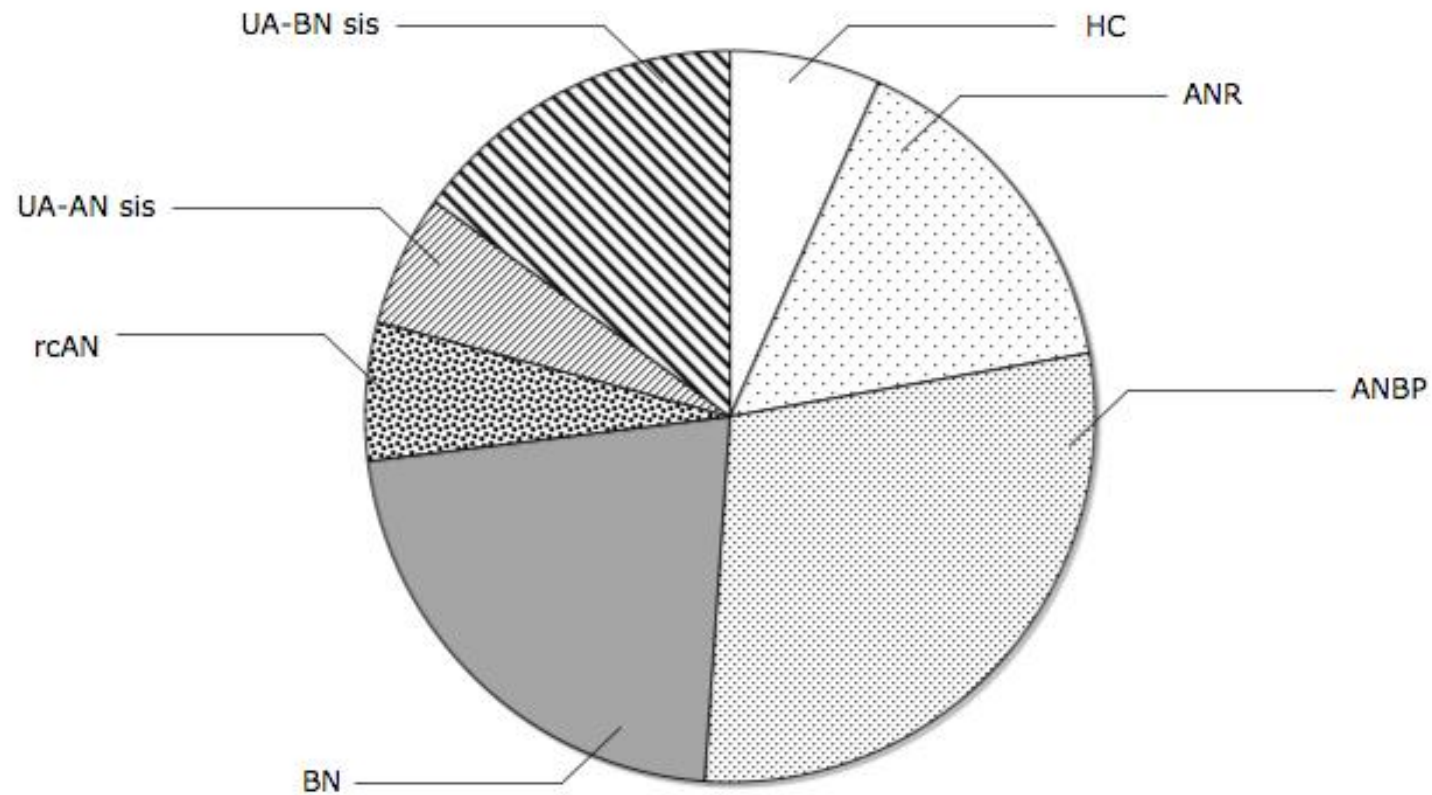
HC Healthy Control; ANR Restricting type Anorexia Nervosa; ANBP Binge/purging type Anorexia Nervosa; BN Bulimia Nervosa; rcAN recovered Anorexia Nervosa; UA-AN sis Unaffected AN sister; UA-BN sis Unaffected BN sister

associated with poor prognostic features. Women with current ED and impaired shifting were more likely to have a longer and more severe illness, more comorbid anxiety diagnoses, lower self-esteem, and were more likely to have engaged in self-harming behaviours. Sample sizes were not large enough to merit this comparison in the recovered or unaffected sister groups.

11.2.2 Weak coherence findings

11.2.2.1 Weak Coherence as an endophenotypes of ED

Weak coherence in women with ED was investigated using a transdiagnostic approach across the first, third and last criteria of an endophenotype. The first hypothesis was that women with current ED would show superior local processing compared to HC women. This hypothesis was confirmed on both tasks, where women with ED were faster on the GEFT (small effect) and had a lower (more detail focussed) coherence index (moderate to large effect) with intact accuracy. The third hypothesis was that sister pairs concordant for an ED would not differ on measures of coherence. This hypothesis was confirmed, with no differences seen between sister pairs concordant for an ED.

The last hypothesis was that unaffected sisters of those with ED would show heightened local processing compared to HC women. This hypothesis was confirmed in part, where a very large effect on the ROCF coherence index (with intact accuracy- small effect) indicated superior local processing. However, a negligible to small difference that was not significant was seen on the GEFT, where unaffected sisters were on average slower than HC indicating a lack of superior local processing. These transdiaganostic findings are not in line with each other, suggesting that weak coherence may have distinct presentations across AN and BN unaffected sisters.

11.2.2.2 Weak Coherence as an endophenotype of AN

Weak coherence in AN was investigated across the first, second and last criteria outlined above. The first hypothesis was that women with current AN would a bias toward local or detailed processing compared to HC women. This hypothesis was confirmed across both tasks. Those with current AN were significantly faster than HC on the GEFT (moderate effect) and had a significantly lower coherence index on the ROCF (moderate effect) with intact accuracy (negligible effect) compared to HC. This finding was consistent for restricting type AN, however GEFT results did not reach significance (small effect size) in those with a mixture of AN

and BN behaviours. This study was the second replication of superior local processing in AN on the EFT or GEFT (Tokley & Kemps, 2007; Lopez et al., 2008b) where the memory component of the task has been removed, and the first replication using the coherence index on the ROCF (Lopez et al., 2008b). A similar finding of enhanced local processing in AN has been found on the Matching Familiar Figures Test (Southgate, Tchanturia, & Treasure, 2008b). Likewise, a systematic review of superior local processing in ED shows this trait has also been found on measures such as the block design, object assembly and fragmented pictures tasks (Lopez et al., 2008c), where moderate effects are seen.

The second hypothesis was that women recovered from AN would not differ on measures of coherence compared to current AN, but would show heightened local processing compared to HC women. This hypothesis was confirmed in that women recovered from AN employed local processing on the GEFT where it assisted task performance (negligible difference to current AN, moderate difference to HC), however like HC employed a more global strategy on the ROCF (moderate difference to current AN, negligible difference to HC). Thus those recovered from AN still had heightened local processing, however they were able to adopt a different strategy when it was not advantageous. This finding is in line with the notion that the endophenotype may require challenge or provocation when assessed in a recovered population (Walters & Owen, 2007). One previous study has investigated coherence in a mixed recovered cohort, where detail focus persisted across both the EFT and ROCF tasks. (Lopez et al., 2008e). Differences in the clinical composition in both groups could explain this difference, however given this is only the second study of weak coherence in ED replication is required to further understand the mechanism involved in recovered cohorts.

The last hypothesis was that unaffected sisters of those with AN would show heightened local processing compared to HC women. This hypothesis was confirmed, where unaffected AN sisters showed consistent detail focus across both the GEFT (small to moderate effect) and the ROCF (very large effect) with accuracy scores on the ROCF comparable to HC (negligible effect size). Effect sizes were comparable if not larger to those seen in the direct AN/HC analysis in Study 1. Differences between sisters were negligible to small. This was the first study to investigate weak coherence in discordant sister pairs, thus providing the first empirical evidence for weak coherence meeting criteria 4 of an endophenotype in

AN. Similar investigations of unaffected siblings and parents in the autistic spectrum population have revealed smaller effects that are mixed across studies (Fombonne et al., 1997; Briskman, Happe, & Frith, 2001; Bolte & Poustka, 2006; Delorme et al., 2007).

Overall, these findings provided strong evidence for the superior local processing (detail focus) aspect of weak coherence as an endophenotype of AN.

11.2.2.3 Weak Coherence as an endophenotype of BN

Weak coherence in BN was investigated across the first and last criteria of an endophenotype. The first hypothesis was that women with current BN would show a bias toward local or detailed processing compared to HC women. This hypothesis was not confirmed. On the ROCF, a significantly lower coherence index (large effect size) and significantly lower accuracy score (moderate effect size) compared to HC were rather suggestive of poor global integration. The lack of a local bias in BN was supported by GEFT results, where time taken was comparable to HC (negligible effect size). The same pattern was seen when those with BN and no history of AN were analysed separately. While the current findings for BN on the ROCF are the same as those obtained previously (Lopez et al., 2008d), results on the GEFT differ. This may in part be due to faster performance here than that found in previous HC groups. Replication including a clean measure of global integration is required in order to understand whether a trade-off exists between detail focus and global integration in the BN population.

Given the findings in Study 1, the last hypothesis was that unaffected sisters of those with BN would show poor global integration compared to HC women. This hypothesis was confirmed, in that the same pattern was seen in unaffected sisters as for women with BN on both the ROCF (large effect for coherence index and accuracy) and the GEFT (negligible effect). This was the first investigation of weak coherence in unaffected sisters or relatives of women with BN.

Overall, these findings provided strong evidence for the poor global integration aspect of weak coherence as an endophenotype of BN. Investigation of this trait in women recovered from BN is still required.

11.2.2.4 Incidence and impact of weak coherence on illness factors

The two coherence measures did not correlate significantly, indicating that the tasks were not measuring a unified concept. Therefore results could not be collapsed across tasks as was the case with the set-shifting data. Rather, results were

collapsed across strategy (good or poor performance on each task) creating four quadrants of coherence; adaptive detail focus, persistent detail focus, persistent global focus and a maladaptive strategy. Persistent detail focus (low coherence index and fast GEFT time) was the most endorsed quadrant, with just under half of all current ED cases (regardless of subtype) employing this strategy. Figure 33 depicts the weighted proportion of persistent detail focus cases across participant groups. Detailed strategies were most common in women with ANR, where 70% of cases showed either adaptive or persistent detail focus. Over half of the women recovered from AN showed adaptive detail focus, with none exhibiting maladaptive detail focus. Persistent detail focus was also most common in unaffected sisters of those with AN, however a maladaptive strategy was most endorsed by unaffected BN sisters. These combined results are in line with those presented across tasks.

Current ED cases showing either adaptive or persistent detail focus were selected for further analysis, to investigate their relationship with clinical variables. Persistent detail focus was associated with poor prognostic factors, such as moderately higher self-report depression and anxiety, and a significantly higher likelihood of social phobia, specific phobia, multiple anxiety diagnoses and alcohol dependence. As with set-shifting, sample sizes were not large enough to merit this comparison in those recovered from AN or the unaffected sisters of those with ED.

11.2.3 The relationship between set-shifting and weak coherence

The relationship between intact/impaired shifting and coherence quadrants was explored in women with lifetime ED. Those with adaptive detail focus were most likely to have intact shifting (90%). Approximately one third of those with persistent detail or persistent global focus had impaired shifting, while over half of those with maladaptive detail focus showed impaired shifting. A small proportion of women with a lifetime ED diagnosis presented with both impaired shifting and persistent detail focus (15.3%; see Figure 34, Figure 35 and Figure 36). This was least common in those with ANR. Those with both traits were more likely to have a more severe ED as indicated by a lower BMI and more severe YBC preoccupations and rituals than those with either trait in isolation. This was the first investigation of the relationship between set-shifting and coherence in ED.

Figure 33: Weighted proportion of persistent detail focus cases across participants groups
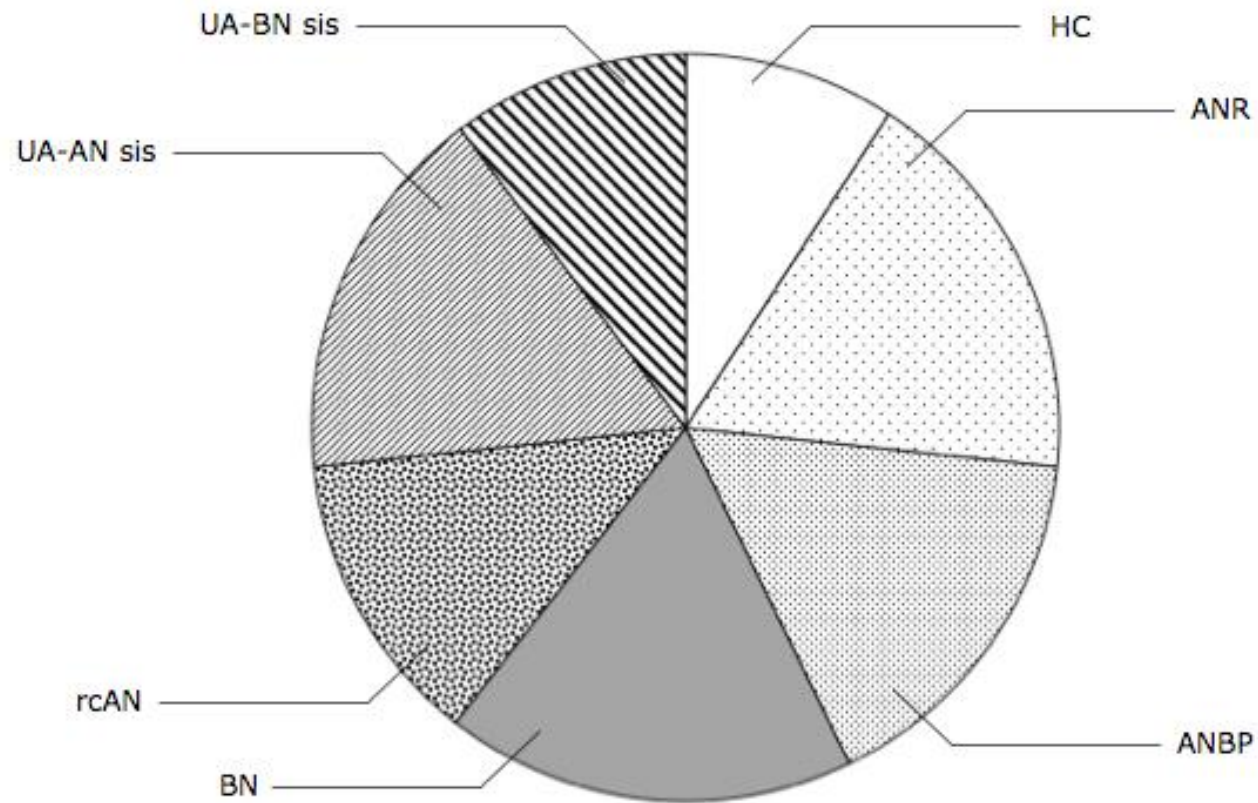
HC Healthy Control; ANR Restricting type Anorexia Nervosa; ANBP Binge/purging type Anorexia Nervosa; BN Bulimia Nervosa; rcAN recovered Anorexia Nervosa; UA-AN sis Unaffected AN sister; UA-BN sis Unaffected BN sister

Figure 34: Illustration of the incidence of impaired shifting, persistent detail focus, and simultaneous presentation of both traits in women with lifetime ED.
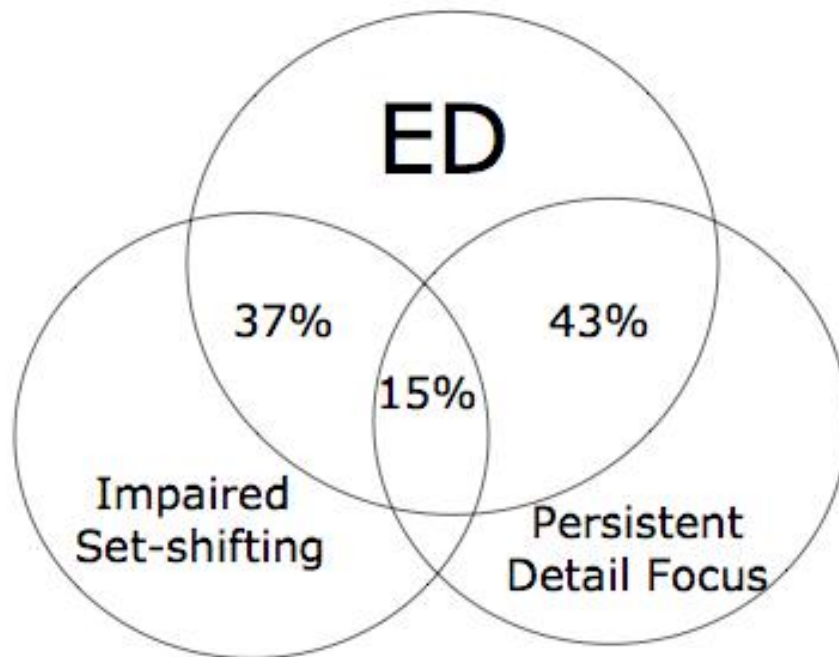

Figure 35: Illustration of the incidence of impaired shifting, persistent detail focus, and simultaneous presentation of both traits in women with current AN
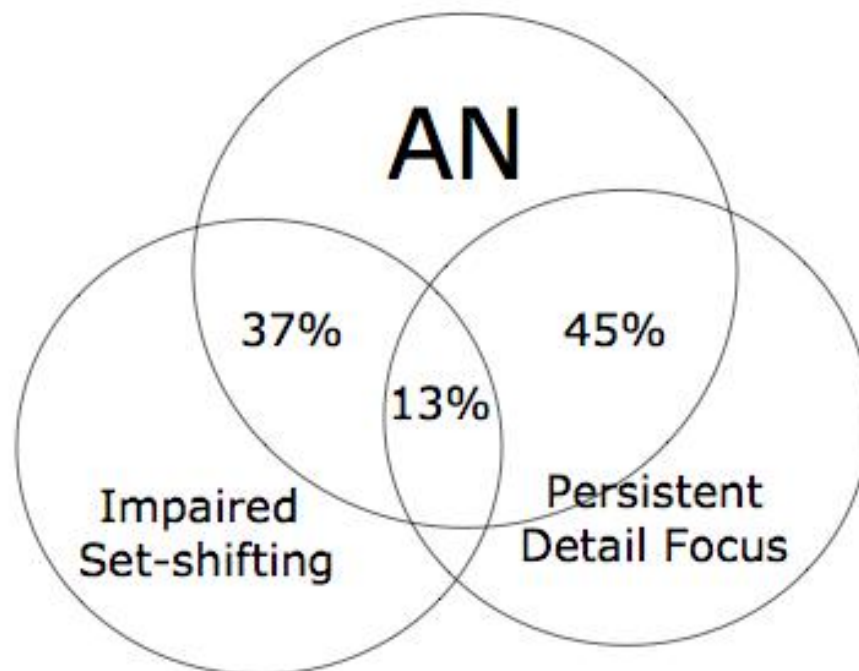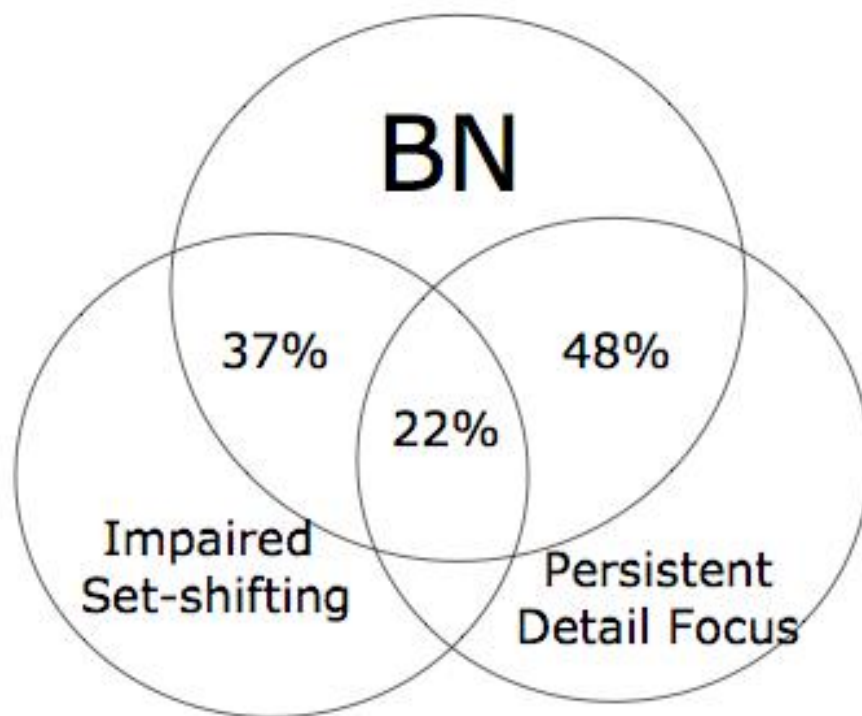
Figure 36: Illustration of the incidence of impaired shifting, persistent detail focus, and simultaneous presentation of both traits in women with current BN

11.2.4 Secondary findings

Two self-report measures of flexibility, the Cognitive Flexibility Scale (CFS), and the Thinking Styles Questionnaire (TSQ) were employed in the current study. Both measures showed an endophenotype pattern across the four empirical studies with moderate to very large effect sizes. However the CFS and TSQ correlated better with self-report depression, anxiety and self-esteem than with neuropsychological outcomes, suggesting that they were perhaps more a measure of an individual's self-evaluation in the context of their psychological state than a measure of cognitive style. It is not recommended that current self-report measures of cognitive style be employed in place of neuropsychological testing in the ED population.

In terms of comorbidity, lifetime depression was the most endorsed in the current ED group (75%), followed by OCD at just under 50% of the sample. In the recovered AN population, the same pattern for lifetime comorbidity was observed with endorsement rates only minimally lower. Unaffected sisters of those with ED were also most likely to endorse depression, with approximately one third of sisters meeting lifetime criteria. OCD was not as common as social phobia (16% threshold) in unaffected sisters. This finding differs slightly to a previous study of psychiatric illness in unaffected sisters of AN, where depression also the most highly endorsed (20%) followed by panic disorder (15%), where only 2% (n=1) met criteria for social phobia (Karwautz, Rabe-Hesketh, Collier, & Treasure, 2002).

11.3 Strengths of the design

This thesis was explicitly designed around the criteria of an endophenotype, and is the first investigation of neuropsychological traits as candidate endophenotypes of ED where all criteria as outlined by Gottesman and Gould (2003) have been simultaneously addressed. Not only is this study design unique in investigating endophenotypes of ED, it is also unique with regards to the more general field of psychiatric endophenotypes. Recent systematic reviews have synthesised evidence across one or more criteria, for example in bipolar (Bora et al., 2008), and schizophrenia (Snitz et al., 2006). Despite their usefulness, these reviews provide only a partial illustration of neurocognitive traits as candidate endophenotypes, given their narrow focus (e.g. neurocognition in unaffected 1st degree relatives only). While a large number of theoretical papers outlining the advantages of the endophenotype approach are found in the ED literature (Bulik et

al., 2007b; Steiger & Bruce, 2007; Treasure et al., 2007; Treasure, 2007), substantial empirical research is as yet sparse. Only a handful of papers exist where the main aim was to address one or more endophenotype criteria (Holliday et al., 2005; Steiger et al., 2006; Lopez et al., 2008e). By investigating all criteria within the same cohort, this thesis has made a significant contribution to the search for endophenotypes of ED, by providing a comprehensive overview of both set-shifting and weak coherence across all aspects of an endophenotype.

The identification of endophenotypes of ED has the potential to lead to a) a more appropriate classification system for ED based on underlying biological causes rather than phenotypic presentation, b) the development of biologically driven models of the development and maintenance of ED, and c) design and implementation of targeted, more effective treatment interventions for ED to replace current treatment largely borrowed from other psychiatric disorders. Results from this thesis help inform each of these three goals.

Firstly, this thesis has found impaired set-shifting and weak coherence in women with ED compared to HC women. It has also highlighted that these traits are not universal in the ED population in that a subset of those with ED have poor set-shifting, another subset have weak coherence, another subset have both traits and yet others neither of the two. These traits occurred across diagnostic categories, indicating that current diagnosis is not a reliable predictor of trait presence. Thus, aspects of the underlying biological mechanism resulting in disordered eating behaviour provide a form of subtyping ED based on endophenotype rather than phenotype. Such an approach to ED classification may be more appropriate than the current phenotypic diagnostic criteria, particularly given the substantial instability of the current DSM-IV ED diagnoses (Anderluh, Tchanturia, Rabe-Hesketh, Collier, & Treasure, 2009), and the large number of patients presenting to community clinics that do not meet these criteria (Fairburn, Cooper, Bohn, O'Connor, Doll, & Palmer, 2007). Coherence and set-shifting ability may form two aspects of underlying ED traits that could contribute to a more biologically based classification system.

Secondly, results from this thesis inform theoretical models of ED by providing empirical evidence for neurocognitive anomalies contributing to the biological basis and maintenance of ED. Specifically, one component of Schmidt and Treasure's (2006) maintenance model of AN is that of obsessive-compulsive personality traits (of which cognitive rigidity and detail focus are hypothesised to fall

under) and perfectionism. The model outlines these traits as being present before the individual presents with AN and then being exacerbated by the illness. As this thesis design was cross-sectional rather than longitudinal in nature, results cannot inform this assumption directly. It can however be tentatively proposed that trait presence in unaffected sisters provides support for poor set-shifting and weak coherence being present before the illness onset, and a higher level of trait presence in current compared to recovered AN provides support for them being exacerbated by the illness. With regard to future models of ED, results from this thesis may contribute the first empirical data toward an endophenotypic model of the development of ED. With the identification of further endophenotypes of ED, such a model has the potential to predict clinical characteristics, prognosis and response to treatment based on a range of underlying biological mechanisms. Such a model would be empirically driven and largely free from cultural influences.

Thirdly, this thesis has provided a substantial increase in knowledge with regard to the clinical presentation of those with and without notable neurocognitive biases. A thorough picture of clinical features and psychiatric comorbidities for 128 women with an ED (and 50 unaffected sisters) has been detailed here. Comorbidity rates for women with ED were in general similar to those found in large-scale genetics projects such as the Price Foundation study (Kaye et al., 2004). In addition to presenting these features by diagnostic group, this thesis has also investigated these features based on the presence or absence of poor set-shifting and/or persistent detail focus. This is the first study known to explore clinical presentation based on neurocognitive profile rather than diagnostic criteria. The finding that women with these traits showed poorer prognostic factors has significant implications for treatment development, where a modularised approach to ED treatment would provide a useful platform from which to integrate these findings. For example, based on a neurocognitive profile of poor set-shifting and adaptive detail processing, a treatment plan could be developed to integrate modules specifically targeting flexibility, anxiety, self-esteem, eating related rituals and self-harm together with a core ED module. In comparison, based on a neurocognitive profile of intact shifting but persistent detail focus, modules targeting local processing bias, anxiety, depression, social cognition and alcohol use could be added to the core ED module. Such an approach to treatment, which would also incorporate the clinical correlates of other candidate endophenotypes, allows for a targeted, individualised treatment

plan for each patient where biological mechanisms in addition to life story are taken into account.

11.4 Weaknesses of the design

A number of methodological weaknesses were present in the design of this thesis, largely because of inherent shortcomings in the neuropsychological tasks employed. These weaknesses are outlined by task below, along with suggestions on how these issues might be overcome in future studies employing similar methodology.

11.4.1 CatBat task

A number of methodological considerations should be raised regarding the CatBat task. As detailed in the methodology chapter (see 5.1.1.5), the paragraph consists of a Cat half and a Bat half. Ideally, both halves of the story should be matched in order to detect a difference in time to complete each half (and errors made). While both halves are matched in terms of completions required (6/6) and approximate number of words (66/69), one of the completions in the first half of the story is a suffix phoneme (making a cat completion plural: _AT_), while all other completions are prefix phonemes only (_AT). As this is the only place where a suffix phoneme occurs (i.e. it is not balanced by a second suffix phoneme in the second half of the story), the two halves are unbalanced. In administering the task, it was observed that many participants would noticeably pause at this completion, adding more seconds to their Cat time. Secondly, during the design of the current protocols the author was not aware that there were two versions of the CatBat task- one that finished with "and the _at *rushed away*" and another that finished "and the _at *flew away*". This difference may seem minimal however the *rushed away* sentence makes logical sense with either a bat or a cat completion, whereas *flew away* makes sense only for a bat completion. The *rushed away* version was unintentionally selected for the current study, having an impact both on errors and on Bat time (therefore the balanced B-C variable). As this completion was the last in the story, falling in the Bat section, it was decided that a cat completion in this space would consistently be marked as an error. However this oversight also had an impact on time, in that a large number of participants again would noticeably pause on this final completion, deciding which of the two plausible responses they would choose thus elongating their BAT time.

It could be argued that one of these errors adds time to the CAT half and the other to the BAT half, thus their combined impact may balance out. However it is suggested that the effects of these errors cannot be so simply explained given that each posed a unique situation. Further administration of this task should type in the required plural (making the final Cat completion _ats) and use the *flew away* completion at the end of the BAT half of the story. Alternatively, given that a notable number of participants were unfamiliar with the general behaviour of a Bat (e.g. uncertain whether a bat screams or not), it may be pertinent to apply the same test construction to a story paragraph where the key words are both well known in the general population (e.g. Cat/Rat, or Dog/Log).

11.4.2 Trail Making Test (TMT)

In addition to the CatBat task, the design of the TMT task should be mentioned here. Across all three TMT trials (Baseline, Alphabet, Alphanumeric) the sequence in which dots were to be connected in the test trial was identical: Regardless of whether dots were connected using the labels A-B-C-D-E etc or 1-A-2-B-3 etc, the physical sequence in which dots on the screen were selected remained the same. This means that an element of implicit learning could have improved performance on the later trials (e.g. shifting trial), which may have impacted on the sensitivity of the task. Future research using the TMT should design a different physical dot sequence across the three trials, controlling for overall sequence 'travel' distance (mouse movement) across the three trials.

11.4.3 Group Embedded Figure Test (GEFT)

Both the CatBat and GEFT tasks required the assessor to use a stopwatch in order to measure outcome. In the case of the GEFT, the stopwatch must be re-started after a false claim. It is possible that an element of experimenter bias emerged, depending on how fast the stopwatch was re-started after the false claim. Additionally, in the case of a participant having a very fast response time, the outcome could be more a factor of the experimenter's reaction time than the participant's. Problems occurred with a particular stopwatch that did not respond well to the stop/start button, making some results inaccurate. In the later stages of data collection, a touch-screen stopwatch was used, where timing accuracy was perceived as much improved. In order to ensure accuracy, a touch-screen digital stopwatch or a computerised version of this task would be ideal.

While all clinical participants completed the GEFT using the blue shaded booklets, approximately half of the HC sample were assessed with a black and white copy of the booklet. It is possible that shapes in the black and white version of the blue shaded booklets were easier to segment, contributing to the faster reaction time of the HC group. Although such small details can seem frivolous, future studies should ensure that identical stimuli are administered to all participants in order to avoid any methodological differences impacting the results.

11.4.4 Brixton task

The main concern regarding the Brixton task is that of order effects. The Brixton task followed the WCST in the neuropsychological battery. This decision was initially made as instructions for the WCST did not alert the participant to a potential change in pattern throughout the task, however instructions for the Brixton did. Therefore it was required that the WCST be administered before the Brixton task, so that a prior task instruction to look for a pattern change did not influence results. The tasks followed each other simply because they were both computerised (along with the TMT), and the battery order was designed to complete all computerised tasks first, followed by the pen and paper tasks in order to avoid unnecessary time spend setting up and putting away the laptop computer. In retrospect, administering these three flexibility tasks in a row may have served as flexibility practise, priming the participant to perform more flexibly on each subsequent task. Therefore regardless of an explicit task instruction regarding shifting, the observation of less errors than previously found in clinical groups on the Brixton task (the last of the three computerised shifting tasks) may have been influenced by order effects, which contributed to the sensitivity of the task. Future studies should pay careful attention to the order of task batteries, and where possible ensure that tasks tapping the same construct do not follow each other directly.

11.5 Limitations of the current study

Limitations pertinent to individual studies were presented at the end of each appropriate chapter (see chapters 6-9). The main considerations of note are outlined below, along with details of the endeavours made to overcome them or recommendations for future studies where this was not possible.

11.5.1 Healthy control sample considerations

Throughout this thesis, a consistent theme of unexpected HC performance on a number of the neuropsychological tasks has been raised. As all of the employed tasks have been utilised in a number of previous studies both within our unit and a number of collaborators, we have gained a good understanding of the expected performance of a control group. While in a few cases (Haptic, GEFT) results were still in line with the hypotheses, the margin of difference between clinical and control groups was substantially smaller than that found in previous research. Three reasons are proposed for these differences. Firstly, experimenter effects could have played a role. While all clinical participants were assessed by the author, HC data was collected by three different masters level psychologists (LD, BW, EB). These assessors were given extensive training and supervision regarding neuropsychological assessment specific to the tasks employed here. However, tasks where HC results are under question (CatBat, Haptic, GEFT) are tasks administered in the traditional face-to-face fashion where the experimenter is actively involved in task administration, rather than having a more passive role as in the case of computerised tasks or simple picture copying. Thus, a difference in style across assessors may have confounded results, whether it be experimenter reaction time when using the stopwatch in the case of the CatBat or GEFT tasks, or employing a slower administration style in the case of the Haptic task (thus making participants over-analyse the size disparity).

Secondly, HC participants may have been too extensively screened. The majority of willing HC were screened out because of family history of a mental health problem, or due to scoring above the cut-off on self-report measures. Given the focus of the current study on endophenotypes, it was important that those with a 1st degree relative with a mental health illness were screened out, particularly as impaired set-shifting is found across a number of psychiatric groups such as schizophrenia and may therefore be present in the unaffected relatives of these clinical groups. However, the current criteria also meant that a cousin or aunt with mental health problems would lead to a participant being excluded. Given that a UK population survey found 21% of the general population to suffer from mental illness (NationalStatistics, 2003), excluding HC based on any known relative with mental illness could have created a limited and unrepresentative control sample. Additionally, cut-off scores for self-report measures could have been less

conservative, for example using a score of 10 for HADS cut-off (probable anxiety/depression) rather than a score of eight (possible anxiety/depression) (Zigmond & Snaith, 1983).

Finally, as a major recruitment drive for HC was conducted through King's College London student circular, a student sample bias could have been present. Due to the campus our unit is on, nearly all respondents in the initial phase (assessed by LD) were medical students, and a large proportion in the third phase (assessed by EB) were postgraduate psychology students. While a concerted effort was made in the other stages of HC recruitment to balance this student sample by targeting a broad range of female participants (e.g. mothers & business women), initial use of cluster and convenience sampling may have impacted the current findings.

11.5.1.1 Efforts made to overcome limitation

It was possible to assess one of these potential confounders formally. Following initial analysis of the HC data (n=65; LD=20, BW=21, EB=24), the author (MR) recruited and assessed a subset of HC participants equivalent in size to that collected by each of the three student assessors (n=23). Comparisons were made across tasks and assessors to investigate experimenter effects, i.e. whether results differed between the subset of HC assessed by the author compared to student assessors. Investigation of boxplots revealed notably different distributions across assessors despite the lack of significant findings (see Appendix 9 for boxplots). In general, a tighter range was observed for LD (medical student population) compared to the wider range for BW (general community population). Despite this, analysis of the data using ANOVA and Kruskal-Wallis tests revealed no significant between-assessor differences. Effect sizes between HC assessors for each neuropsychological task were also calculated. Moderate effects were found between assessors on the TMT (0.40; LD/MR), Brixton (0.69; LD/EB), CatBat (0.44; BW/MR), Haptic (0.58; EB/BW) and GEFT (0.43; EB/BW). A small effect size was found on the WCST (0.34; BW/MR) and a negligible effect size on the ROCF (0.11; EB/MR). All tasks where HC results are under question (CatBat, Haptic, GEFT) showed moderate effect sizes, while the fully automated WCST and the ROCF (where MR scored all of the drawings) showed little difference between groups. Thus, effect size analysis suggests that a substantial difference in HC results was evident for some tasks depending on which HC assessor conducted the assessment. It is difficult to conclude

whether the variance between HC groups is indicative of experimenter or sampling effects.

Overall, analysis between HC samples suggests that the difference between the current combined HC group and that of previous samples on the CatBat, Haptic and GEFT tasks may be in part due to experimenter or sampling effects. Ideally, future studies should employ one consistent assessor and match HC to clinical participants on a one-to-one basis using demographic variables such as gender, age, IQ, socio-economic status and/or area of residence (postcode/borough/county). The difficulties of such a matching procedure in a large-scale project are acknowledged, however such a strategy may be the only way to be confident of clinical and control groups being equivalent. Asking clinical participants to suggest an unrelated friend or colleague as a HC is time-efficient way of recruiting an approximate demographic match.

11.5.2 General methodological considerations

Perhaps the most notable limitation of the current study is that a measure of general intelligence was not employed. A measure of intelligence (IQ) could have been useful both in order to match clinical and control groups, and/or to run as a covariate on neuropsychological tasks thought to be influenced by IQ. Two of the tasks employed here are of specific concern; the GEFT and WCST. The Embedded Figures Test has been found to correlate negatively with verbal intelligence in HC but not ED populations (Lopez et al., 2008b). The WCST has been found to be related to fluid intelligence however for the total errors variable rather than perseverative errors as used here (Salthouse, Atkinson, & Berish, 2003). Therefore, while it is unlikely that the lack of an IQ measure substantially affected the current findings, future studies of neuropsychological profile should ensure that an approximate measure of IQ is included. Ideally, such a measure should be employed to match clinical and control participants. In addition to IQ, information regarding current medication was not initially collected from participants. This data was collected during the neuropsychological assessment session for approximately the last 20% of the sample. Finally, as discussed in Study 2 (see 7.5.1), the limitation of a cross-sectional design to assess the state-independence criteria of an endophenotype is acknowledged.

11.5.2.1 Efforts made to overcome limitation

Retrospective collection of a measure of IQ was deemed unfeasible. Data on whether clinical participants were taking psychotropic medication was retrospectively collected by email and telephone. Although every effort was made to reach every individual, not all participants could be reached or complied with the request. Current medication details were collected from 60% of the clinical sample. While no effect of current medication on neuropsychological performance was found in the current study, the high number of missing cases may have affected this outcome. Previous studies of both set-shifting and coherence in ED have not found effects of medication on neuropsychological performance (Holliday et al., 2005; Lopez et al., 2008b).

11.6 Treatment application

There is significant potential for the application of work on neuropsychological profile to the treatment of ED. Treatment modules incorporating our understanding of cognition in ED (such as those detailed in this theses) are already underway. The development and implementation of both inpatient and outpatient pre-treatment modules will be outlined in brief below.

11.6.1 Inpatient application

11.6.1.1 Background

The application of neuropsychological findings in ED to inpatient treatment has been under development within our unit at the Institute of Psychiatry over the past 5 years (Tchanturia et al., 2007; Tchanturia, Davies, Lopez, Schmidt, Treasure, & Wykes, 2008). Cognitive remediation therapy (CRT) is focussed on addressing the underlying *process* rather than *content* of cognition, through the use of simple cognitive exercises to enhance reflection, meta-cognition and therefore increase awareness of cognitive style. It is hypothesised that CRT works by both encouraging brain connectivity in the formation of new neural circuitry, in addition to providing the patient with more adaptive cognitive strategies (Tchanturia et al., 2007). CRT for ED was developed from a similar module for schizophrenia to make it applicable to the ED population. Longitudinal research on CRT in schizophrenia has found improvements in cognitive strategies (e.g. flexibility), social functioning and illness symptomology (McGurk, Twamley, Sitzer, McHugo, & Mueser, 2007).

11.6.1.2 Structure and content of CRT for ED

CRT consists of 10, 45-minute sessions delivered weekly or twice weekly on a one-to-one basis by a trained CRT therapist. This therapist could be from a wide range of backgrounds such as psychology, nursing or social work. CRT in its current form is designed to address both rigid and detail focussed thinking styles in addition to perfectionism as a pre-therapy to more complex psychological intervention. Using over 15 different types of cognitive exercises (each with multiple variations), the patient and therapist work collaboratively to reflect on and explore the strategies the patient employs to complete each task, therefore enhancing the patient's meta-cognition (ability to think about their thinking). See Figure 37 for one of the illusion tasks, an example of an exercise in CRT where the patient must shift their focus to see the two different women in the picture. Other flexibility tasks include token towers (patient and therapist stack different sized, coloured and shaped tokens on a pile according to a changing rule determined by therapist or patient e.g. blue/small/triangle) and the traditional stroop task. Tasks designed to tap detail/global processing are the geometric figures task (patient verbally describes a shape for the therapist to draw) and the main idea task (patient summarises the gist of a written letter/article). Once familiar with the session content, it is possible to work through upward of 10 exercises per session

While CRT is a manualised intervention (Tchanturia & Davis, In Progress), the goals across sessions are similar and flexibility exists with regard to which exercises are chosen in a given session. Each session has in common the following four components; reflection and discussion of thinking style employed, exploration of the pros and cons of that thinking style in the current context and in other situations, learning and practising more adaptive (flexible/global) strategies, and completion of behavioural experiments between sessions designed to promote the development of these new strategies. Emphasis placed on each component alters as the patient moves through the sessions. At first, development of the therapeutic alliance and reflection skills are prioritised, while sessions toward the end of the intervention focus on practising adaptive strategies through the use of behavioural experiments. The final session of CRT ends with both patient and therapist exchanging 'good-bye letters'. See Table 62 for an outline of the priorities of CRT across sessions, and further details regarding the good-bye letters.

Figure 37: Example of an illusion task used in CRT (old woman & young lady)

Table 62: CRT priorities across sessions (Tchanturia & Davis, In Progress)

Sessions	Main components
1-3	- Building up a collaborative therapeutic alliance - Explaining the rationale of CRT for AN - Introducing and practicing exercises to identify the predominant cognitive style - Encouraging of making links between cognitive exercises and behaviour out of session
4-6	- Mainly practicing cognitive exercises - Reflection of strengths and weaknesses of predominant cognitive style - Designing behavioural experiments in session - Practicing behavioural experiments between sessions - Reflecting on the results and strategies learnt in the behavioural experiments and how to overcome obstacles - Encourage transfer of skills to daily life
6-8	- Practicing cognitive exercises - Greater emphasis on designing, practicing and discussing behavioural experiments than in earlier sessions - Encouraging of making links between behavioural experiments and behaviours in real life - Preparing for the end of CRT
9	- Same as sessions 6-8 - Reflecting on and discussing strategies to maintain changes after CRT - Reflecting on and discussing difficulties that might arise after CRT and how they could be overcome - Introducing 'good bye letter' exchange for next session: A motivational strategy where both patient and therapist summarise and reflect on the experience of CRT, cognitive styles, new strategies learnt, main achievements, areas that need further reinforcement, maintenance of changes, and provide some guidance on overcoming possible future obstacles.
10	- Exchanging and discussing good bye letters - Ending CRT

11.6.1.3 Empirical evidence for CRT

Empirical evidence to date for CRT is largely in the form of qualitative reports and case series'. Early qualitative feedback from patients has been used to modify and improve the content of CRT (Davies & Tchanturia, 2005; Tchanturia et al., 2007). A case series following four in-patients with AN found moderate to large effects of improvement on neuropsychological measures of set-shifting (TMT, Brixton, CatBat, Haptic) following CRT (Tchanturia et al., 2007). A subsequent study delivered an amended form of CRT (including both set-shifting and global processing components) to 23 inpatients with AN (Tchanturia et al., 2008). Significant improvements in BMI (baseline 14.1, follow-up 16.1), and most neuropsychological tasks were seen, along with a significant drop in self-report depression levels. A small drop-out rate was observed (14.8%). These findings show promise for the effectiveness of a cognitive intervention as a pre-therapy module for acute AN. Given the design of the investigations to date, it is difficult to parse out practise effects as confounder given that the same assessment tasks were used at both baseline and follow-up. A randomised trial of CRT amongst in-patients will help to address this limitation, in addition to aiding identification of the additional benefit of CRT over treatment as usual.

11.6.2 Outpatient application

11.6.2.1 Background and development

More recently, an outpatient form of CRT has begun development. The goal of increasing flexibility and global processing remains consistent. However given the higher weight and therefore increased cognitive functioning of outpatients with AN, the design of the intervention differs markedly. Inpatient CRT conducts a neuropsychological assessment before and after the intervention, however makes no direct reference to the assessment results themselves. In contrast, 'neuro-feedbackgives the patient direct feedback on their assessment results, which forms the basis of the short intervention.

Neuro-feedback is based on motivational enhancement therapy (MET). MET employs the therapeutic style of motivational interviewing (MI), whilst also incorporating an element of personalised feedback into therapy. This could include direct feedback on the patients health (e.g. for smoking or alcohol cessation) or neuropsychological feedback as is the case here. MI is a directive, client-centred style of interaction designed to bring about behaviour change in a non-judgemental

and collaborative fashion (Miller & Rollnick, 2002). This is done by exploring and resolving ambivalence toward change within the so-called "spirit of MI", which includes collaboration, evocation, and autonomy. The therapist uses four key principals to encourage this tone: expressing empathy, developing discrepancy, rolling with resistance, and supporting self-efficacy.

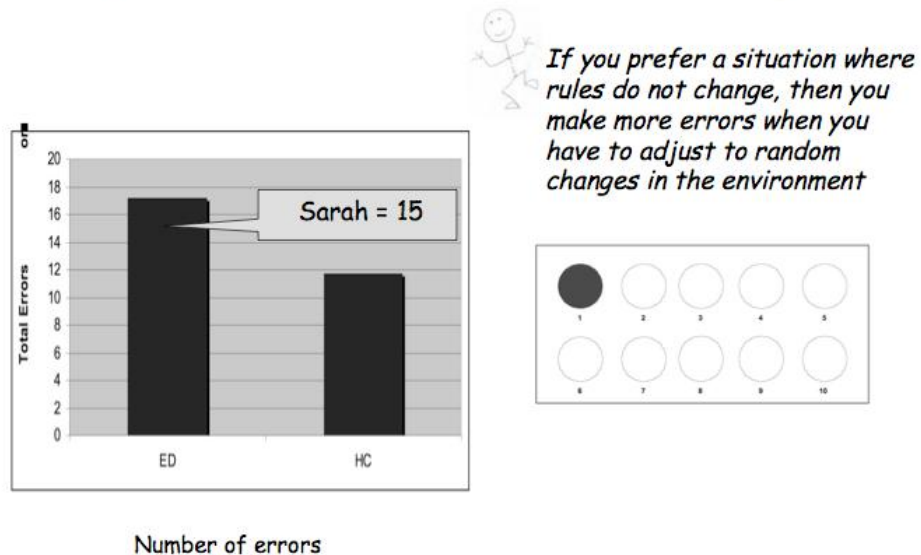### 11.6.2.2 Structure and content of outpatient feedback for ED

In it's current form, neuro-feedback consists of a short, three-session module spread over two or three weeks. Sessions cover 1) neuropsychological assessment, 2) feedback, formulation and target setting, and 3) reflection. The initial session consisting of the neuropsychological assessment, is often done in conjunction with a general psychiatric assessment. In the second session, the patient is given the results from their neuropsychological assessment in the form of personalised visual feedback with clinical and control norms. Figure 38 shows an example of two feedback slides for 'Sarah'. One can see that Sarah is in the normal range for someone with an ED on both tasks, and is even slightly faster than usual on the GEFT. Using MI, the therapist works through feedback slides from each task, creating a formulation as to the impact (pros and cons) of the patient's traits as evidenced by the neuropsychological assessment on aspects of everyday life (relationships, academic/occupational work) in addition to how they shape eating behaviour. Target setting lends from cognitive behavioural therapy, where a goal is formulated and behavioural experiments are designed to help the patient transcend their natural bias or adopt a less extreme position in a given situation.

### 11.6.2.3 Empirical evidence for outpatient feedback

To date, one published paper outlines the neuro-feedback intervention, leaning on two cases as illustrations (Lopez, Roberts, Tchanturia, & Treasure, 2008a). One case (with good prognostic features) shows substantial improvement directly following the intervention, describing the feedback as a turning point for her. Both cases comment positively on their experience of neuro-feedback, stating that it helped them understand their own thinking better, which in turn helped them to understand aspects of the development and maintenance of their ED. A further case series of neuro-feedback with seven patients with an ED (5 AN; 1 EDNOS; 1 BED) outlines similar positive qualitative reports, where improvements in BMI (drop in BMI for the BED patient) are seen both directly after the intervention and at 2 month

Figure 38: Example slides from Neuro-feedback for Brixton (top) and GEFT (bottom) tasks
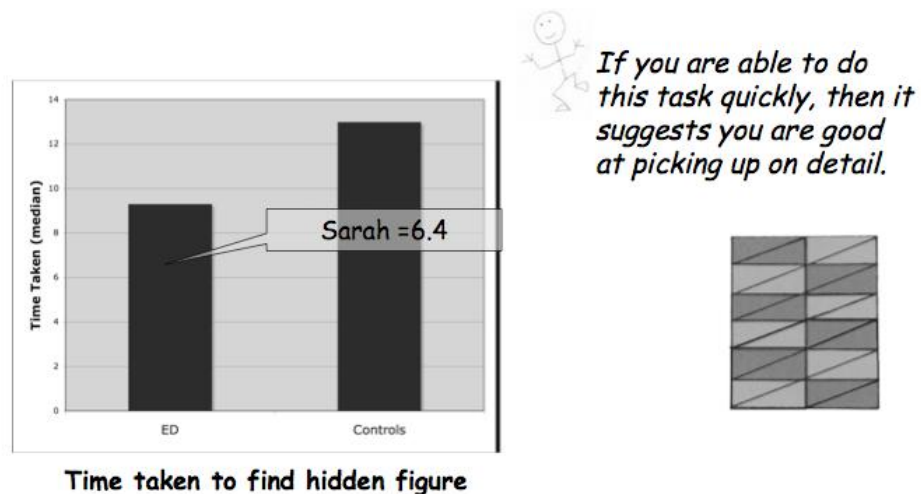
follow-up (Lopez, 2008). These early reports suggest that neuro-feedback, like CRT, may be a useful pre-therapy module where the patient can 1) gain some insight into their own cognitive biases and the impact of these on life and their ED, and 2) form a therapeutic alliance and trust in the process of therapy within the non-threatening context of thinking style, before moving onto more complex psychological intervention targeting their ED behaviours.

11.6.3 Application to family work

Evidence for the presence of poor set-shifting and weak coherence in unaffected sisters of those with an ED has been provided in this thesis (see 9.2). It may therefore be appropriate to incorporate some form of neuro-feedback into family therapy for ED. Feedback of cognitive style across the family (siblings and parents,in addition to the individual with ED) may help diffuse blame and guilt that is often present within families with an ED. It may also provide something for family members to share and relate to each other on, for example if the ED sister, an unaffected sibling and a parent all share the trait of cognitive rigidity. Additionally, family feedback may provide a platform for family members to encourage the patient to be more flexible or more globally focussed on a daily basis, essentially bringing aspects of the therapy into the home environment.

11.6.4 Practical considerations for clinical application

It is of clinical interest to examine the presence of these neuropsychological traits in patients and how they shape behaviour. Implementation of CRT and/or neuro-feedback in general clinical practise requires the presence of a trained neuropsychological assessor to administer, score, and provide feedback to patients. For the majority of clinical units (particularly those in the public health sector), resource constraints mean that expensive, complex and time-consuming neuropsychological assessment is not practical. However a general idea as to the neuropsychological profile of a patient would assist clinical formulation. In such cases, it would be of use to have a small battery of simple tasks that were fast to administer, score and interpret. Where time and resources are particularly stretched, administration of the Brixton task (as a measure of set-shifting ability) and the GEFT (as a measure of detail focus) is recommended. It may be possible to administer just one half of the GEFT to participants and retain task sensitivity, in order to save on administration time (Davis, Harrison et al., In Progress). One must however be cautious when drawing conclusions on the presence or absence of a cognitive trait

based on one task alone. Where possible, it is suggested that two measures of each concept are administered; the Brixton task and WCST as measures of set-shifting, and the GEFT and either the ROCF or Fragmented Pictures Task (Snodgrass, Smith, Feenan, & Corwin, 1987), as measures of coherence. While the ROCF would be the author's recommendation, the time required for an individual to be trained in the scoring methodology for both accuracy and coherence index makes this task unfeasible for most clinical settings. The fragmented pictures task has recently been trialed in our unit with a large effect size of 1.1 found between 40 current AN and 42 HC participants (Harrison, Personal Communication). This task requires participants to identify a gradually appearing object as fast as possible (e.g. elephant, book). It is a more pure measure of global integration than the ROCF and thus compliments the GEFT well. To ensure that order effects do not cloud task effects, it is suggested that the four tasks be administered in the following order: 1) ROCF/fragmented pictures task, 2) WCST, 3) GEFT, 4) Brixton task.

11.7 Future directions

Future studies following this line of work will play a key role in further understanding aspects of neurocognition as endophenotypes of ED. In addition to the methodological improvements suggested above (see 11.4 and 11.5), six main considerations for future studies are outlined below.

11.7.1 Replication and expansion of the current study

Following from the current thesis, it would be pertinent to increase the statistical power of the results presented here by adding to the numbers of participants. This is particularly relevant to the BN group, where only 17 individuals with BN and no history of AN were included, and only 20 sister pairs discordant for (lifetime) BN were able to be collected. Power analysis indicated that between 20-40 participants per group were required to detect differences across neuropsychological tasks. Therefore a larger sample size across all clinical groups but particularly pure BN and their unaffected sisters may help clarify the trends observed in this thesis.

11.7.2 Molecular genetics

The analysis of genetic data in relation to neurocognitive performance is an essential step in determining whether set-shifting and weak coherence are biomarkers or heritable endophenotypes. Understanding of the genetic basis of ED lags behind the research base of other psychiatric conditions. A limited number of linkage and

association studies are available in the ED literature, with most findings hindered by small sample sizes and lack of replication (Bulik, Slof-Op't Landt, van Furth, & Sullivan, 2007c). Three linkage studies to date have implicated a number of candidate genes in relation to the phenotypes of AN and BN (rather than diagnostic categories) for example lowest ever BMI, drive for thinness, and food-related obsessionality (Devlin et al., 2002; Grice et al., 2002; Bacanu et al., 2005). Linkage has been found with regions on chromosomes 1, 3 and 13, along with specific neurotransmitter genes such as the serotonin D1 receptor HTRD1, and the opoid delta receptor OPRD1 (Bergen et al., 2003). Association studies have largely focussed on serotonergic and dopaminergic genes. Some specific single nucleotide polymorphisms (SNPs) have been identified, for example within the serotonin D1 receptor gene and the dopamine D2 receptor and COMPT genes (for a review, see Bulik et al., 2007c). Genome wide association studies, which allow investigation of markers or SNPs across complete sets of DNA (the human genome) will significantly advance our understanding of genetic variation in those with ED and are currently underway. More studies with larger sample sizes are required, with particular attention paid to the collection of lifetime phenotypic presentation in order to map genotype with illness phenotypes and endophenotypes.

Blood samples collected from this study's participants require DNA extraction and analysis. These samples will form part of a genome wide association study of AN. Following the findings of this thesis, analysis of genetic data for both this sample and that of further familial studies of neurocogntion in ED should be made a priority.

11.7.3 Neurobiological investigations

In addition to genetic analysis, understanding the neurobiological underpinnings of set-shifting and weak coherence in ED is a crucial next step for this area of research. Recent advances in the tools to measure brain structure (MRI), blood flow in response to stimuli (fMRI), and specific neurotransmitter systems (PET, SPECT) will substantially increase knowledge of the neurobiology of ED both generally and in relation to neurocognitive functioning. Early imaging studies in AN used PET to investigate disturbances in serotonin and dopamine pathways in small samples of recovered AN patients (Frank et al., 2002; Frank et al., 2005). It is thought that disturbances in serotonin contribute to appetite dysregulation, anxiety, obsessionality and extreme impulse control, while alterations of the dopamine

system contribute to altered reward, decision making and executive control (Kaye, 2008). More recently, researchers have begun to realise the potential of stimulus driven investigations. For example, altered brain activation in the insula and in frontal and temporal regions has been found in relation to body image words (Redgrave et al., 2008), while body related images produce altered somatosensory processing in AN (Santel, Baving, Krauel, Munte, & Rotte, 2006). FMRI studies have found that women recovered from AN show an altered insula response to food stimuli compared to control women (Wagner et al., 2008), and have altered reward processing as demonstrated by greater activation in the caudate (Wagner et al., 2007). One recent study has used fMRI to investigate neural activation in women with BN when performing the Simon Spatial Incompatibility Task (Marsh et al., 2009). Findings indicated that a failure to control frontostriatal circuits during the task resulted in poorer self-regulation in BN.

As an aspect of executive control, cognitive flexibility is largely regulated by the prefrontal cortex, a region rich in the neurochemical systems of dopamine and serotonin (Robbins, 2000). This area is particularly sensitive to changes in dopamine levels. FMRI investigations in the general population have implicated brain regions/structures such as the dorsolateral prefrontal cortex (Ravizza & Carter, 2008) and the caudate nuclei (Graham et al., 2008) along with altered dopamine in the prefrontal cortex as impacting on set-shifting ability (Roberts et al., 1994; Nagano-Saito, Leyton, Monchi, Goldberg, He, & Dagher, 2008). Research in schizophrenia has found a relationship between prefrontal atrophy and poor set-shifting (Bonilha et al., 2008). One study to date has investigated the neural correlates of poor set-shifting in AN, where increased activation in frontoparietal regions during task completion is observed (Zastrow et al., 2009). It seems likely that poor set-shifting seen in those with ED may share some relationship with altered dopamine levels in the prefrontal cortex. Also implicated in brain plasticity is brain-derived neurotropic factor (BDNF), a protein that mediates appetite regulation through the serotonergic system. Serum BDNF concentrations are lower in AN compared to control women (Nakazato et al., 2003; Monteleone, Fabrazzo, Martiadis, Serritella, Pannuto, & Maj, 2005) and are no different in weight recovered AN (Nakazato, Hashimoto, Yoshimura, Hashimoto, Shimizu, & Iyo, 2006). A relationship between low serum BDNF and poor cognitive flexibility has been found in schizophrenia using a stroop task and the TMT (Han et al., 2008). One study has correlated serum BDNF with

WCST results in the AN population, however no relationship was found using the percentage (rather than raw) perseverative errors variable (Nakazato et al., 2008).

Understanding the neural basis of weak coherence has made some progress in the autism literature. Two main theories exist (Happe & Frith, 2006). The first proposes atypical functioning in specific brain pathways or regions, such as a deficit in the dorsal visual pathway (Spencer, O'Brien, Riggs, Braddick, Atkinson, & Wattam-Bell, 2000), the magnocellular pathway (Milne, Swettenham, Hansen, Campbell, Jeffries, & Plaisted, 2002), or anomalies throughout the right hemisphere which may impair global processing (McKelvey, Lambert, Mottron, & Shevell, 1995; Waiter, Williams, Murray, Gilchrist, Perrett, & Whiten, 2005). The second proposes generally reduced neural connectivity throughout the brain (Brock, Brown, Boucher, & Rippon, 2002; Just, Cherkassky, Keller, & Minshew, 2004). Using fMRI, a pilot study found that some but not all areas of activation are similar across autistic and control participants while completing the EFT (Ring et al., 1999). Normal controls showed greater activity in prefrontal and parietal areas (suggesting use of executive functioning/working memory), while those with autism instead showed greater activation in occipital regions, suggesting a heavier reliance on mental imagery (heightened visual assessment of the element) to complete the task. No published studies could be found where imaging has been used to investigate coherence in the ED population.

The costly and time-consuming nature of scanning requires a systematic methodological approach, ideally employed across a multi-centre collaboration. Pilot trials of simple set-shifting and coherence tasks being completed in the scanner are currently underway within our department, and are ripe for future research.

11.7.4 Further family studies

In addition to replication of the current sister pair design, it would be of interest to examine these same neurocognitive traits in parents of those with an ED. Such studies could examine differences between women with an ED where no parent, one parent or both parents share these traits. Such investigations would not only help inform the search for endophenotypes but could also indicate whether any compounding effect is present (in the case of both parents being rigid or detail focussed), or whether these traits genetically filter through fathers or mothers. Mixed evidence is found for detail focus presenting more strongly in mothers or fathers of those with autism (Happe et al., 2001; Bolte & Poustka, 2006). In the schizophrenia

literature mothers and fathers have not been separately investigated, however parents show poor set-shifting compared to controls (Dollfus et al., 2002; Appels, Sitskoorn, Westers, Lems, & Kahn, 2003; Ma et al., 2007). The cognitive profile of parents of those with ED is yet to be investigated.

Also of importance would be the examination of these traits in twins that are both concordant and discordant for an ED. Discordant twin studies provide a unique opportunity to examine disorder specific mechanisms, as the unaffected twin can act as a perfect genetic comparison. Twin study designs in ED are already being used to explore candidate endophenotypes such as temperament (Wilksch & Wade, 2008). In such investigations, careful administration of lifetime clinical diagnostic interviews is required in order to differentiate twins with current and lifetime full syndrome, partial syndrome and no ED behaviours. When assessing neurocognitive traits, full Axis I comorbidity should also be assessed given that such neurocognitive anomalies are present across psychiatric conditions.

11.7.5 Other aspects of cognitive functioning

While this thesis has focussed on two aspects of neurocognition, there remain other aspects of executive and cognitive functioning that are yet to be examined as candidate endophenotypes. Work in our unit has already begun on emotional processing (Harrison, Sullivan, Tchanturia, & Treasure, Submitted). Impulsivity and reward sensitivity have also been implicated to a degree (Treasure et al., 2007), however to date these traits have only been investigated in light of the first endophenotype criteria. Although it was not a focus, this thesis has highlighted low self-esteem as a putative endophenotype as evidenced by lower ratings in current ED, recovered AN, and unaffected ED sisters compared to HC. High perfectionism may also fit this pattern. These findings provide support for the use of self-report measures to investigate psychological endophenotypes of ED. In terms of process, investigations into new candidate endophenotypes should start with an assessment of the candidate trait in the current and recovered phases of the illness. Given the difficulties recruiting adequate numbers of sister pairs into research (both those concordant and discordant for an ED), it would be prudent to have at least two candidate traits with evidence of state-independence before approaching this scarce sample. Careful attention should be paid to the selection of neuropsychological tasks and self-report measures in these emerging areas, preferably involving initial trials of a number of tasks tapping the same concept (e.g. impulsivity) in order to investigate

the validity and sensitivity of new measures in the ED population. This approach has been strategic in the past, to inform selection of both set-shifting and coherence tasks.

11.7.6 Longitudinal research

Finally, longitudinal research in this area is required. This thesis raised the limitation of a cross-sectional design in assessing state-independence. Using this methodology, it is impossible to conclude whether those recovered from the illness also 'recovered' a more normative style of cognitive functioning, or whether they were part of the minority of those with an ED who did not endorse these traits even when they were ill (which may have contributed to recovery). Employing a longitudinal design is the most reliable way to understand the predictive value of neurocognition in recovery from ED.

11.8 Concluding comment

This thesis has provided moderate evidence for poor set-shifting meeting four of the five criteria for a candidate endophenotype of ED in a sample of 270 women with and without lifetime ED. It has also provided strong evidence for aspects of weak coherence as candidate endophenotypes of ED, specifically persistent detail focus in AN and evidence of poor global integration in BN. These traits are not rare in the ED population, with approximately 70% of current ED cases exhibiting one or both traits. A bias toward cognitive rigidity and persistent detail focus are associated with poor prognostic features such as comorbid anxiety diagnoses. These traits are not specific to the ED population, and are likely general endophenotypes of psychiatric illness. The identification of psychiatric endophenotypes has the potential to significantly advance the ED field by informing biologically based classification systems, development/maintenance models, and approaches to treatment. Treatment modules targeting these traits show promising results, indicating that identification and integration of cognitive traits into treatment may result in a more positive outcome for individuals with ED.

12   References

Ackard, D. M., Fulkerson, J. A., & Neumark-Sztainer, D. (2007). Prevalence and utility of DSM-IV eating disorder diagnostic criteria among youth. *International Journal of Eating Disorders, 40*(5), 409-417.

Anderluh, M., Tchanturia, K., Rabe-Hesketh, S., Collier, D., & Treasure, J. (2009). Lifetime course of eating disorders: design and validity testing of a new strategy to define the eating disorders phenotype. *Psychological Medicine, 39*(1), 105-114.

APA. (2000). *Diagnostic and Statistical Manual of Mental Disorders-Revised (DSM-IV-TR)*. Washington DC: American Psychiatric Association.

Appels, M. C., Sitskoorn, M. M., Westers, P., Lems, E., & Kahn, R. S. (2003). Cognitive dysfunctions in parents of schizophrenic patients parallel the deficits found in patients. *Schizophr Res, 63*(3), 285-293.

ArmyIndividualTestBattery. (1944). *Manual of directions and scoring*. Washington D.C.: War Department, Adjutant General's Office.

Bacanu, S. A., Bulik, C. M., Klump, K. L., Fichter, M. M., Halmi, K. A., Keel, P., Kaplan, A. S., Mitchell, J. E., Rotondo, A., Strober, M., Treasure, J., Woodside, D. B., Sonpar, V. A., Xie, W., Bergen, A. W., Berrettini, W. H., Kaye, W. H., & Devlin, B. (2005). Linkage analysis of anorexia and bulimia nervosa cohorts using selected behavioral phenotypes as quantitative traits or covariates. *Am J Med Genet B Neuropsychiatr Genet, 139B*(1), 61-68.

Ball, J., & Mitchell, P. (2004). A randomized controlled study of cognitive behaviour therapy and behavioural family therapy for anorexia nervosa patients. *Eating Disorders, 12*, 303-314.

Bardone-Cone, A. M., Wonderlich, S. A., Frost, R. O., Bulik, C. M., Mitchell, J. E., Uppala, S., & Simonich, H. (2007). Perfectionism and eating disorders: current status and future directions. *Clinical Psychology Review, 27*(3), 384-405.

Barrett, S. L., Kelly, C., Bell, R., & King, D. J. (2008). Gender influences the detection of spatial working memory deficits in bipolar disorder. *Bipolar Disord, 10*(5), 647-654.

Bayless, J. D., Kanz, J. E., Moser, D. J., McDowell, B. D., Bowers, W. A., Andersen, A., & Paulsen, J. S. (2002). Neuropsychological characteristics of patients in a hospital-based eating disorder program. *Annals of Clinical Psychiatry, 14*(4), 203-207.

Becker, K. A. (2003). *History of the Stanford-Binet intelligence scales: Content and psychometrics (Stanford-Binet Intelligence Scales, Fifth Edition Assessment Service Bulletin No. 1)*. Itasca, IL: Riverside Publishing.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1988-1101.

Bergen, A. W., van den Bree, M. B., Yeager, M., Welch, R., Ganjei, J. K., Haque, K., Bacanu, S., Berrettini, W. H., Grice, D. E., Goldman, D., Bulik, C. M., Klump, K., Fichter, M., Halmi, K., Kaplan, A., Strober, M., Treasure, J., Woodside, B., & Kaye, W. H. (2003). Candidate genes for anorexia nervosa in the 1p33-36 linkage region: serotonin 1D and delta opioid receptor loci exhibit significant association to anorexia nervosa. *Mol Psychiatry, 8*(4), 397-406.

Bolte, S., & Poustka, F. (2006). The broader cognitive phenotype of autism in parents: how specific is the tendency for local processing and executive dysfunction? *Journal of Child Psychology and Psychiatry, 47*(6), 639-645.

Bonilha, L., Molnar, C., Horner, M. D., Anderson, B., Forster, L., George, M. S., & Nahas, Z. (2008). Neurocognitive deficits and prefrontal cortical atrophy in patients with schizophrenia. *Schizophr Res, 101*(1-3), 142-151.

Booth, R. (2006). *Local-global processing and cognitive style in autism spectrum disorders and typical development.* Kings College University of London, London.

Booth, R., Charlton, R., Hughes, C., & Happe, F. (Eds.). (2003). *Disentangling weak coherence and executive dysfunction: Planning drawing in autism and attention-deficit/hyperactivity disorder.* New York, NY: Oxford University Press.

Bora, E., Yucel, M., & Pantelis, C. (2008). Cognitive endophenotypes of bipolar disorder: A meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. *Journal of Affective Disorders.*

Bradburn, M. J., Deeks, J. J., & Altman, D. G. (1988). sbe24:metan- an alternative meta-analysis command. *Stata Technical Bulletin, 44*, 1-15.

Brambilla, F., Monteleone, P., Bortolotti, F., Dalle Grave, R., Todisco, P., Favaro, A., Santonastaso, P., Ramacciotti, C., Paoli, R., & Maj, M. (2003). Persistent amenorrhoea in weight-recovered anorexics: psychological and biological aspects. *Psychiatry Research, 118*(3), 249-257.

Brecelj-Anderluh, M., Tchanturia, K., Rabe-Hesketh, S., & Treasure, J. (2003). Childhood obsessive-compulsive personality traits in adult women with eating disorders: Defining a broader eating disorder phenotype. *American Journal of Psychiatry, 160*(2), 242-247.

Briskman, J., Happe, F., & Frith, U. (2001). Exploring the cognitive phenotype of autism: Weak "central coherence" in parents and siblings of children in autism: II. Real-life skills and preferences. *Journal of Child Psychology and Psychiatry, 42*(3), 309-316.

Brock, J., Brown, C. C., Boucher, J., & Rippon, G. (2002). The temporal binding deficit hypothesis of autism. *Developmental Psychopathology, 14*(2), 209-224.

Brownley, K. A., Berkman, N. D., Sedway, J. A., Lohr, K. N., & Bulik, C. M. (2007). Binge eating disorder treatment: A systematic review of randomized controlled trials. *International Journal of Eating Disorders, 40*(4), 337-348.

Bulik, C. M., Berkman, N. D., Brownley, K. A., Sedway, J. A., & Lohr, K. N. (2007a). Anorexia nervosa treatment: A systematic review of randomized controlled trials. *International Journal of Eating Disorders, 40*(4), 310-320.

Bulik, C. M., Hebebrand, J., Keski-Rahkonen, A., Klump, K. L., Reichborn-Kjennerud, T., Mazzeo, S. E., & Wade, T. D. (2007b). Genetic epidemiology, endophenotypes, and eating disorder classification. *International Journal of Eating Disorders, 40 Suppl*, S52-60.

Bulik, C. M., Slof-Op't Landt, M. C., van Furth, E. F., & Sullivan, P. F. (2007c). The genetics of anorexia nervosa. *Annu Rev Nutr, 27*, 263-275.

Bulik, C. M., Sullivan, P. F., Tozzi, F., Furberg, H., Lichtenstein, P., & Pedersen, N. L. (2006). Prevalence, Heritability, and Prospective Risk Factors for Anorexia Nervosa. *Archives of General Psychiatry, 63*(3), 305-312.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests.* UK: Thames Valley Test Company Ltd.

Button, E. J., Benson, E., Nollett, C., & Palmer, R. L. (2005). Don't forget EDNOS (eating disorder not otherwise specified): Patterns of service use in an eating disorders service. *Psychiatric Bulletin, 29*(4), 134-136.

Chamay-Weber, C., Narring, F., & Michaud, P.-A. (2005). Partial eating disorders among adolescents: A review. *Journal of Adolescent Health, 37*(5), 417-427.

Chamberlain, S. R., Fineberg, N. A., Blackwell, A. D., Robbins, T. W., & Sahakian, B. J. (2006). Motor Inhibition and Cognitive Flexibility in Obsessive-Compulsive Disorder and Trichotillomania. *American Journal of Psychiatry, 163*(7), 1282-1284.

Channon, S., de Silva, P., Hemsley, D., & Perkins, R. (1989). A controlled trial of cognitive-behavioural and behavioural treatment of anorexia nervosa. *Behaviour Research and Therapy, 27*(5), 529-535.

Clark, L., Sarna, A., & Goodwin, G. M. (2005). Impairment of executive function but not memory in first-degree relatives of patients with bipolar I disorder and in euthymic patients with unipolar depression. *American Journal of Psychiatry, 162*(10), 1980-1982.

Collier, D., & Treasure, J. (2004). The aetiology of eating disorders. *British Journal of Psychiatry, 185*(5), 363-365.

Cooper, H., & Hedges, L. (1994). *The Handbook of Research Synthesis*. New York: The Russel Sage Foundation.

Crisp, A., Gowers, S., Joughin, N., McClelland, L., Rooney, B., Nielsen, S., Bowyer, C., Halek, C., Hartman, D., Tattersall, M., Hugo, P., Robinson, D., Atkinson, R., & Clifton, A. (2006). Anorexia Nervosa in Males: Similarities and Differences to Anorexia Nervosa in Females. *European Eating Disorders Review, 14*(3), 163-167.

Crow, S. J., Agras, S. W., Halmi, K., Mitchell, J. E., & Kraemer, H. C. (2002). Full syndromal versus subthreshold anorexia nervosa, bulimia nervosa, and binge eating disorder: a multicenter study. *International Journal of Eating Disorders, 32*(3), 309-318.

Currin, L., Schmidt, U., Treasure, J., & Jick, H. (2005). Time trends in eating disorder incidence. *British Journal of Psychiatry, 186*(2), 132-135.

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science, 264*(5162), 1102-1105.

Dare, C., Eisler, I., Russell, G., Treasure, J., & Dodge, L. (2001). Psychological therapies for adults with anorexia nervosa: randomised controlled trial of out-patient treatments.[see comment]. *British Journal of Psychiatry, 178*, 216-221.

Davies, H., & Tchanturia, K. (2005). Cognitive remediation therapy as an intervention for acute anorexia nervosa: A case report. *European Eating Disorders Review, 13*(5), 311-316.

Delorme, R., Gousse, V., Roy, I., Trandafir, A., Mathieu, F., Mouren-Simeoni, M. C., Betancur, C., & Leboyer, M. (2007). Shared executive dysfunctions in unaffected relatives of patients with autism and obsessive-compulsive disorder. *European Psychiatry, 22*(1), 32-38.

Devlin, B., Bacanu, S. A., Klump, K. L., Bulik, C. M., Fichter, M. M., Halmi, K. A., Kaplan, A. S., Strober, M., Treasure, J., Woodside, D. B., Berrettini, W. H., & Kaye, W. H. (2002). Linkage analysis of anorexia nervosa incorporating behavioral covariates. *Hum Mol Genet, 11*(6), 689-696.

Dohnt, H., & Tiggemann, M. (2006). The Contribution of Peer and Media Influences to the Development of Body Satisfaction and Self-Esteem in Young Girls: A Prospective Study. *Developmental Psychology, 42*(5), 929-936.

Dollfus, S., Lombardo, C., Benali, K., Halbecq, I., Abadie, P., Marie, R. M., & Brazo, P. (2002). Executive/attentional cognitive functions in schizophrenic patients and their parents: a preliminary study. *Schizophr Res, 53*(1-2), 93-99.

Downes, J., Roberts, A., Sahakian, B., Evenden, J., Morris, R. G., & Robbins, T. W. (1989). Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: Evidence for a specific attentional dysfunction. *Neuropsychologia, 27*(11-12), 1329-1343.

Doyle, A. E. (2006). Executive functions in attention-deficit/hyperactivity disorder. *Journal of Clinical Psychiatry, 67 Suppl 8*, 21-26.

Dunn, E. C., Larimer, M. E., & Neighbors, C. (2002). Alcohol and drug-related negative consequences in college students with bulimia nervosa and binge eating disorder. *International Journal of Eating Disorders, 32*, 171-178.

Egan, M. F., Goldberg, T. E., Gscheidle, T., Weirich, M., Rawlings, R., Hyde, T. M., Bigelow, L., & Weinberger, D. R. (2001). Relative risk for cognitive impairments in siblings of patients with schizophrenia. *Biological Psychiatry, 50*(2), 98-107.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta analysis detected by a simple, graphical test. *British Medical Journal, 315*(629-634).

Eisler, I., Dare, C., Hodes, M., Russell, G., Dodge, E., & Le Grange, D. (2000). Family therapy for adolescent anorexia nervosa: the results of a controlled comparison of two family interventions. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 41*(6), 727-736.

Eliava, N. (1964). *A Problem of Set in Cognitive Psychology*: Tbilisi.

Eslinger, P. J., & Grattan, L. M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia, 31*(1), 17-28.

Everitt, B. S. (2003). *Modern Medical Statistics*. London: Arnold Publishers.

Fairburn, C. G., Cooper, Z., Bohn, K., O'Connor, M. E., Doll, H. A., & Palmer, R. L. (2007). The severity and status of eating disorder NOS: implications for DSM-V. *Behaviour Research and Therapy, 45*(8), 1705-1715.

Faris, P. L., Raymond, N. C., De Zwaan, M., Howard, L. A., Eckert, E. D., & Mitchell, J. E. (1992). Nociceptive, but not tactile, thresholds are elevated in bulimia nervosa. *Biological Psychiatry, 32*(5), 462-466.

Fassino, S., Piero, A., Daga, G. A., Leombruni, P., Mortara, P., & Rovera, G. G. (2002). Attentional biases and frontal functioning in anorexia nervosa. *International Journal of Eating Disorders, 31*(3), 274-283.

Favaro, A., Santonastaso, P., Monteleone, P., Bellodi, L., Mauri, M., Rotondo, A., Erzegovesi, S., & Maj, M. (2007). Self-injurious behavior and attempted suicide in purging bulimia nervosa: Associations with psychiatric comorbidity. *Journal of Affective Disorders*.

Ferraro, F. R., Wonderlich, S., & Jocie, Z. (1997). Performance variability as a new theoretical mechanism regarding eating disorders and cognitive processing. *International Journal of Eating Disorders, 53*, 117-121.

Flint, J., & Munafo, M. R. (2007). The endophenotype concept in psychiatric genetics. *Psychological Medicine, 37*(2), 163-180.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Complusive Inventory:

Development and validation of a short version. *Psychological Assessment, 14*(4), 485-495.

Fombonne, E., Bolton, P., Prior, J., Jordan, H., & Rutter, M. (1997). A family study of autism: Cognitive patterns and levels in parents and siblings. *Journal of Child Psychology and Psychiatry, 38*(6), 667-683.

Fossati, P., Ergis, A. M., & Allilaire, J. F. (2002). Executive functioning in unipolar depression: a review. *Encephale, 28*(2), 97-107.

Fowler, L., Blackwell, A., Jaffa, A., Palmer, R., & Robbins, T. W. (2005). Profile of neurocognitive impairments associated with female in-patients with anorexia nervosa. *Psychological Medicine, 36*, 517-527.

Fox, C. F. (1981). Neuropsycholgocail correlates of anorexia nervosa. *International Journal of Psychiatry in Medicine, 11*, 285-290.

Frank, G. K., Bailer, U. F., Henry, S. E., Drevets, W., Meltzer, C. C., Price, J. C., Mathis, C. A., Wagner, A., Hoge, J., Ziolko, S., Barbarich-Marsteller, N., Weissfeld, L., & Kaye, W. H. (2005). Increased dopamine D2/D3 receptor binding after recovery from anorexia nervosa measured by positron emission tomography and [11c]raclopride. *Biol Psychiatry, 58*(11), 908-912.

Frank, G. K., Kaye, W. H., Meltzer, C. C., Price, J. C., Greer, P., McConaha, C., & Skovira, K. (2002). Reduced 5-HT2A receptor binding after recovery from anorexia nervosa. *Biol Psychiatry, 52*(9), 896-906.

Frantom, L. V., Allen, D. N., & Cross, C. L. (2008). Neurocognitive endophenotypes for bipolar disorder. *Bipolar Disorders, 10*(3), 387-399.

Frieling, H., Roschke, B., Kornhuber, J., Wilhelm, J., Romer, K. D., Guruss, B., Donsch, D., Hillemacher, T., de Zwann, M., Jacoby, G. E., & Bleich, S. (2005). Cognitive impairment and its association with homocystine plasma levels in females with eating disorders- findings from the HEzD study. *Journal of Neural Transmission, 112*, 1591-1598.

Frith, U. (1989). *Autism: Explaining the Enigma*. Oxford: Basil Blackwell.

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research, 14*(5), 449-468.

Gadalla, T., & Piran, N. (2007). Co-occurance of eating disorders and alcohol use disorders in women: A meta analysis. *Archives of Women's Mental Health, 10*(4), 133-140.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The eating attitudes test: psychometric features and clinical correlates. *Psychological Medicine, 12*(4), 871-878.

Geist, R., Heinmaa, M., Stephens, D., Davis, R., & Katzman, D. K. (2000). Comparison of family therapy and family group psychoeducation in adolescents with anorexia nervosa. *The Canadian Journal of Psychiatry / La Revue canadienne de psychiatrie, 45*(2), 173-178.

Gillberg, I., Gillberg, C., Rastam, M., & Johansson, M. (1996). The cognitive profile of anorexia nervosa: A comparative study including a community-based sample. *Comprehensive Psychiatry, 37*(1), 23-30.

Godart, N., Berthoz, S., Rein, Z., Perdereau, F., Lang, F., Venisse, J. L., Halfon, O., Bizouard, P., Loas, G., Corcos, M., Jeammet, P., Flament, M., & Curt, F. (2006). Does the frequency of anxiety and depressive disorders differ between diagnostic subtypes of anorexia nervosa and bulimia? *International Journal of Eating Disorders, 39*(8), 772-778.

Godart, N. T., Perdereau, F., Rein, Z., Berthoz, S., Wallier, J., Jeammet, P., & Flament, M. F. (2007). Comorbidity studies of eating disorders and mood

disorders. Critical review of the literature. *Journal of Affective Disorders, 97*(1-3), 37-49.

Goldberg, T. E., Torrey, E. F., Gold, J. M., Bigelow, L. B., Ragland, R. D., Taylor, E., & Weinberger, D. R. (1995). Genetic risk of neuropsychological impairment in schizophrenia: a study of monozygotic twins discordant and concordant for the disorder. *Schizophrenia Research, 17*(1), 77-84.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Delgado, P., Heninger, G. R., & Charney, D. S. (1989a). The Yale-Brown Obsessive Compulsive Scale: II. Validity. *Archives of General Psychiatry, 46*(11), 1012-1016.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., Heninger, G. R., & Charney, D. S. (1989b). The Yale-Brown Obsessive Compulsive Scale: I. Development, use, and reliability. *Archives of General Psychiatry, 46*(11), 1006-1011.

Gottesman, II, & Shields, J. (1973). Genetic theorizing and schizophrenia. *British Journal of Psychiatry, 122*(566), 15-30.

Gottesman, I. L., & Gould, T. D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry, 160*, 636-645.

Gottschaldt, K. (1926). Uber den einfluss der erfahrung auf die wahrnehmung vin figur 1, uber den einfluss gehaufter eingragung von figuren auf ibre sichbarkeit in umfassenden konfigurationen. *Psychologische Forschung, 8*, 261-317.

Gowers, S., Norton, K., Halek, C., & Crisp, A. H. (1994). Outcome of outpatient psychotherapy in a random allocation treatment study of anorexia nervosa. *International Journal of Eating Disorders, 15*(2), 165-177.

Gowers, S. G., Clark, A., Roberts, C., Griffiths, A., Edwards, V., Bryan, C., Smethurst, N., Byford, S., & Barrett, B. (2007). Clinical effectiveness of treatments for anorexia nervosa in adolescents: randomised controlled trial. *British Journal of Psychiatry, 191*, 427-435.

Graham, S., Phua, E., Soon, C. S., Oh, T., Au, C., Shuter, B., Wang, S. C., & Yeh, I. B. (2008). Role of medial cortical, hippocampal and striatal interactions during cognitive set-shifting. *Neuroimage*.

Grant, D. A., & Berg, E. A. (1984). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a weigl-type card sorting problem. *Journal of Experimental Psychology, 38*, 404-411.

Grice, D. E., Halmi, K. A., Fichter, M. M., Strober, M., Woodside, D. B., Treasure, J. T., Kaplan, A. S., Magistretti, P. J., Goldman, D., Bulik, C. M., Kaye, W. H., & Berrettini, W. H. (2002). Evidence for a susceptibility gene for anorexia nervosa on chromosome 1. *Am J Hum Genet, 70*(3), 787-792.

Grunwald, M., Ettrich, C., Assmann, B., Dahne, A., Krause, W., Busse, F., & Gertz, H.-J. (2001a). Deficits in haptic perception and right parietal theta power changes in patients with anorexia nervosa before and after weight gain. *International Journal of Eating Disorders, 29*, 417-428.

Grunwald, M., Ettrich, C., Krause, W., Assmann, B., Dahne, A., Weiss, T., & Gertz, H.-J. (2001b). Haptic perception in anorexia nervosa before and after weight gain. *Journal of Clinical and Experimental Neuropsychology, 23*, 520-529.

Gull, W. W. (1868). Anorexia Nervosa. *Lancet, ii*, 171-176.

Hajcak, G., Huppert, J. D., Simons, R. F., & Foa, E. B. (2004). Psychometric properties of the OCI-R in a college sample. *Behaviour Research and Therapy, 42*(1), 115-123.

Halmi, K. A., Agras, W., Crow, S., Mitchell, J., Wilson, G., Bryson, S. W., & Kraemer, H. C. (2005). Predictors of Treatment Acceptance and Completion in Anorexia Nervosa: Implications for Future Study Designs. *Archives of General Psychiatry, 62*(7), 776-781.

Halperin, J. M., Trampush, J. W., Miller, C. J., Marks, D. J., & Newcorn, J. H. (2008). Neuropsychological outcome in adolescents/young adults with childhood ADHD: profiles of persisters, remitters and controls. *Journal of Child Psychology and Psychiatry*.

Han, D. H., Park, D. B., Choi, T. Y., Joo, S. Y., Lee, M. K., Park, B. R., Nishimura, R., Chu, C. C., & Renshaw, P. F. (2008). Effects of brain-derived neurotrophic factor-catecholamine-O-methyltransferase gene interaction on schizophrenic symptoms. *Neuroreport, 19*(11), 1155-1158.

Happe, F., Briskman, J., & Frith, U. (2001). Exploring the cognitive phenotype of autism: Weak "central coherence" in parents and siblings of children with autism: I. Experimental tests. *Journal of Child Psychology and Psychiatry, 42*(3), 299-307.

Happe, F., & Frith, U. (2006). The Weak Coherence Account: Detail-focused Cognitive Style in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 36*(1), 5-25.

Happe, F. G., & Booth, R. D. (2008). The power of the positive: revisiting weak coherence in autism spectrum disorders. *Quarterly Journal of Experimental Psychology, 61*(1), 50-63.

Harlow, J. M. (1848). Passage of an iron rod through the head. *Boston Medical and Surgical Journal, 39*, 389-383 (Republished in Journal of Neuropsychiatry and Clinical Neurosciences, 311:281-283).

Harris, E. C., & Barraclough, B. (1998). Excess mortality of mental disorder. *British Journal of Psychiatry, 173*, 11-53.

Harrison, A., Sullivan, S., Tchanturia, K., & Treasure, J. (Submitted). Attentional Biases to Social Cues in Anorexia Nervosa: Results from a Pictorial Stroop Task.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G., & Curtiss, G. (1993). Wisconsin Cart Sorting Test (Version Computer version 4). Odessa, FL: Psychological Assessment Resources.

Henry, J. D. (2006). A meta-analytic review of Wisconsin card sorting test and verbal fluency performance in obsessive-compultive disorder. *Cognitive Neuropsychiatry, 11*, 156-176.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analysis. *British Medical Journal, 327*, 557-560.

Hoek, H. W., & van Hoeken, D. (2003). Review of the prevelance and incidence of eating disorders. *International Journal of Eating Disorders, 34*, 383-396.

Holliday, J., Tchanturia, K., Landau, S., Collier, D., & Treasure, J. (2005). Is impaired set-shifting an endophenotype of anorexia nervosa? *American Journal of Psychiatry, 162*(12), 2269-2275.

Holtkamp, K., Muller, B., Heussen, N., Remschmidt, H., & Herpertz-Dahlmann, B. (2005). Depression, anxiety, and obsessionality in long-term recovered patients with adolescent-onset anorexia nervosa. *European Child and Adolescent Psychiatry, 14*(2), 106-110.

Jacobi, C., Hayward, C., de Zwaan, M., Kraemer, H. C., & Agras, W. (2004). Coming to Terms With Risk Factors for Eating Disorders: Application of

Risk Terminology and Suggestions for a General Taxonomy. *Psychological Bulletin, 130*(1), 19-65.

Just, M. A., Cherkassky, V. L., Keller, T. A., & Minshew, N. J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain, 127*(Pt 8), 1811-1821.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217-250.

Karwautz, A., Rabe-Hesketh, S., Collier, D. A., & Treasure, J. L. (2002). Pre-morbid psychiatric morbidity, comorbidity and personality in patients with anorexia nervosa compared to their healthy sisters. *European Eating Disorders Review, 10*(4), 255-270.

Kaye, W. (2008). Neurobiology of anorexia and bulimia nervosa. *Physiol Behav, 94*(1), 121-135.

Kaye, W. H., Bulik, C. M., Plotnicov, K., Thornton, L., Devlin, B., Fichter, M. M., Treasure, J., Kaplan, A., Woodside, D. B., Johnson, C. L., Halmi, K., Brandt, H. A., Crawford, S., Mitchell, J. E., Strober, M., Berrettini, W., & Jones, I. (2008). The genetics of anorexia nervosa collaborative study: methods and sample description. *International Journal of Eating Disorders, 41*(4), 289-300.

Kaye, W. H., Bulik, C. M., Thornton, L., Barbarich, N., & Masters, K. (2004). Comorbidity of anxiety disorders with anorexia and bulimia nervosa. *American Journal of Psychiatry, 161*(12), 2215-2221.

Kemps, E., Tiggemann, M., Wade, T., Ben-Tovim, D., & Breyer, R. (2006). Selective working memory deficits in anorexia nervosa. *European Eating Disorders Review, 14*(2), 97-103.

Kingston, J., Szmukler, G., Andrews, D., Tress, B., & Desmond, P. (1996). Neuropsychological and structural brain changes in anorexia nervosa before and after refeering. *Psychological Medicine, 26*, 15-28.

Kitabayashi, Y., Ueda, H., Kashima, A., Okamoto, A., Kooguchi, Y., Narumoto, J., Wada, Y., Yamashita, T., & Fukui, K. (2004). Change of cognitive function following body weight recovery in anorexia nervosa. *Sheshin Igaku, 46*, 243-248.

Klump, K. L., Strober, M., Bulik, C. M., Thornton, L., Johnson, C., Devlin, B., Fichter, M. M., Halmi, K. A., Kaplan, A. S., Woodside, D. B., Crow, S., Mitchell, J., Rotondo, A., Keel, P. K., Berrettini, W. H., Plotnicov, K., Pollice, C., Lilenfeld, L. R., & Kaye, W. H. (2004). Personality characteristics of women before and after recovery from an eating disorder. *Psychological Medicine, 34*(8), 1407-1418.

Koba, T., Shrie, S., & Nabeta, Y. (2002). Impaired performance of wisconsin card sorting test in patients with eating disorders: a preliminary study. *Seishin Igaku 44*, 681-683.

Koskela, A. K., Keski-Rahkonen, A., Sihvola, E., Kauppinen, T., Kaprio, J., Ahonen, A., & Rissanen, A. (2007). Serotonin transporter binding of [123I]ADAM in bulimic women, their healthy twin sisters, and healthy women: a SPET study. *BMC Psychiatry, 7*, 19.

Kravariti, E., Morris, R. G., Rabe-Hesketh, S., Murray, R. M., & Frangou, S. (2003). The Maudsley Early-Onset Schizophrenia Study: cognitive function in adolescent-onset schizophrenia. *Schizophrenia Research, 65*(2-3), 95-103.

Kremen, W. S., Eisen, S. A., Tsuang, M. T., & Lyons, M. J. (2007). Is the Wisconsin Card Sorting Test a useful neurocognitive endophenotype? *Am J Med Genet B Neuropsychiatr Genet, 144B*(4), 403-406.

Kremen, W. S., Faraone, S. V., Seidman, L. J., Pepple, J. R., & Tsuang, M. T. (1998). Neuropsychological risk indicators for schizophrenia: a preliminary study of female relatives of schizophrenic and bipolar probands. *Psychiatry Research, 79*(3), 227-240.

Lauer, C. J., Gorzewski, B., Gerlinghoff, M., Backmund, H., & Zihi, J. (2002). Neuropsychological assessments before and after treatment in patients with anorexia nervosa and bulimia nervosa. *Journal of Psychiatric Research, 33*, 129-138.

Laurent, A., Duly, D., Murry, P., Foussard, N., Boccara, S., Mingat, F., Dalery, J., & d'Amato, T. (2001). WCST performance and schizotypal features in the first-degree relatives of patients with schizophrenia. *Psychiatry Res, 104*(2), 133-144.

Laurent, A., Gilvarry, C., Russel, A., Mathieu-Cura, C., & Murray, R. (2003). Personality dimensions and neuropsychological performance in first-degree relatives of patients with schizophrenia and by affective psychosis. *Encephale, 29*(6), 527-534.

Laurent, A., Moreaud, O., Bosson, J. L., Naegele, B., Boucharlat, J., Saoud, M., Dalery, J., & D'Amato, T. (1999). Neuropsychological functioning among non-psychotic siblings and parents of schizophrenic patients. *Psychiatry Res, 87*(2-3), 147-157.

le Grange, D., Eisler, I., Dare, C., & Russell, G. F. (1992). Evaluation of family treatments in adolescent anorexia nervosa: A pilot study. *International Journal of Eating Disorders, 12*(4), 347-357.

le Grange, D., & Lock, J. (2007). *Treating bulimia in adolescents: A family-based approach.* New York, NY: Guilford Press.

Le Grange, D., & Schmidt, U. (2005). The treatment of adolescents with bulimia nervosa. *Journal of Mental Health, 14*(6), 587-597.

Lena, S. M., Fiocco, A. J., & Leyenaar, J. K. (2004). The role of cognitive deficits in the development of eating disorders. *Neuropsychology Review, 14*(2), 99-113.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological assessment (4th ed.).* New York, NY: Oxford University Press.

Lilenfield, L. R., Kaye, W. H., Greeno, C. G., Merikangas, K. R., Plotnicov, K., Pollice, C., Rao, R., Strober, M., Bulik, C. M., & Nagy, L. (1998). A controlled family study of anorexia nervosa and bulimia nervosa: Psychiatric disorders in first-degree relatives and effects of proband comorbidity. *Archives of General Psychiatry, 55*(7), 603-610.

Lock, J. (2005). Adjusting Cognitive Behavior Therapy For Adolescents With Bulimia Nervosa: Results Of Case Series. *American Journal of Psychotherapy, 59*(3), 267-281.

Lock, J., Agras, W., Bryson, S., & Kraemer, H. C. (2005). A Comparison of Short- and Long-Term Family Therapy for Adolescent Anorexia Nervosa. *Journal of the American Academy of Child & Adolescent Psychiatry, 44*(7), 632-639.

Lock, J., Grange, D. l., Agras, W. S., & Dare, C. (2001). *Treatment manual for anorexia nervosa: A family-based approach.* New York, NY: Guilford Press.

Lopez, C., Roberts, M. E., Tchanturia, K., & Treasure, J. (2008a). Using neuropsychological feedback therapeutically in treatment for anorexia

nervosa: Two illustrative case reports. *European Eating Disorders Review, 16*, 1-10.

Lopez, C., Tchanturia, K., Stahl, D., Happe, F., Booth, R., Holliday, J., & Treasure, J. (2008b). An investigation of central coherence in women with anorexia nervosa. *International Journal of Eating Disorders, 41*(2), 143-152.

Lopez, C., Tchanturia, K., Stahl, D., & Treasure, J. (2008c). Central coherence in eating disorders: A systematic review. *Psychological Medicine, epub ahead of print*.

Lopez, C. A. (2008). *An Examination of Central Coherence in Eating Disorders and its Clinical Implications.* King's College London, London.

Lopez, C. A., Tchanturia, K., Stahl, D., & Treasure, J. (2008d). Central coherence in women with bulimia nervosa. *International Journal of Eating Disorders, 41*(4), 340-347.

Lopez, C. A., Tchanturia, K., Stahl, D., & Treasure, J. (2008e). Weak Central Coherence in Eating Disorders: A step towards looking for an endophenotype of eating disorders. *Journal of Clinical and Experimental Neuropsychology, 4*, 1-9.

Ma, X., Wang, Q., Sham, P. C., Liu, X., Rabe-Hesketh, S., Sun, X., Hu, J., Meng, H., Chen, W., Chen, E. Y., Deng, W., Chan, R. C., Murray, R. M., Collier, D. A., & Li, T. (2007). Neurocognitive deficits in first-episode schizophrenic patients and their first-degree relatives. *Am J Med Genet B Neuropsychiatr Genet, 144B*(4), 407-416.

Marsh, R., Steinglass, J. E., Gerber, A. J., Graziano O'Leary, K., Wang, Z., Murphy, D., Walsh, B. T., & Peterson, B. S. (2009). Deficient activity in the neural systems that mediate self-regulatory control in bulimia nervosa. *Arch Gen Psychiatry, 66*(1), 51-63.

Martin, M., & Rubin, R. (1995). A new measure of cognitive flexibility. *Psychological Reports, 76*, 623-626.

Martin, M. M., & Anderson, C. M. (1998). The Cognitive Flexibility Scale: Three validity studies. *Communication Reports, 11*(1), 1-9.

Masjuan, M. G., Aranda, F. F., & Raich, R. M. (2003). Bulimia nervosa and personality disorders: A review of the literature. *International Journal of Clinical and Health Psychology, 3*(2), 335-349.

Mathers, C. D., Vos, E. T., Stevenson, C. E., & Begg, S. J. (2000). The Australian burden of disease study: Measuring the loss of health from diseases, injuries and risk factors. . *Medical Journal of Australia, 172*, 592-596.

Mathias, J. L., & Kent, P. S. (1998). Neuropsychological consequences of extreme weight loss and dietary restriction in patients with anorexia nervosa. *Journal of Clinical and Experimental Neuropsychology, 20*, 548-564.

Matsunaga, H., Kaye, W. H., McConaha, C., Plotnicov, K., Pollice, C., & Rao, R. (2000). Personality disorders among subjects recovered from eating disorders. *International Journal of Eating Disorders, 27*(3), 353-357.

Mazure, C. M., Halmi, K. A., Sunday, S. R., Romano, S. J., & Einhord, A. M. (1994). The Yale-Brown-Cornell Eating Disorder Scale: Development, use, reliability and validity. *Journal of Psychiatric Research, 28*(5), 425-445.

McGurk, S. R., Twamley, E. W., Sitzer, D. I., McHugo, G. J., & Mueser, K. T. (2007). A meta-analysis of cognitive remediation in schizophrenia. *American Journal of Psychiatry, 164*(12), 1791-1802.

McIntosh, V. V. W., Jordan, J., Carter, F. A., Luty, S. E., McKenzie, J. M., Bulik, C. M., Frampton, C. M. A., & Joyce, P. R. (2005). Three psychotherapies for

anorexia nervosa: a randomized, controlled trial. *American Journal of Psychiatry, 162*(4), 741-747.

McKelvey, J. R., Lambert, R., Mottron, L., & Shevell, M. I. (1995). Right-hemisphere dysfunction in Asperger's syndrome. *J Child Neurol, 10*(4), 310-314.

Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change* (2 ed.). New York: The Guilford Press.

Milne, E., Swettenham, J., Hansen, P., Campbell, R., Jeffries, H., & Plaisted, K. (2002). High motion coherence thresholds in children with autism. *J Child Psychol Psychiatry, 43*(2), 255-263.

Miyake, A., Freidman, N. P., Emerson, M. J., Wizki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Monteleone, P., Fabrazzo, M., Martiadis, V., Serritella, C., Pannuto, M., & Maj, M. (2005). Circulating brain-derived neurotrophic factor is decreased in women with anorexia and bulimia nervosa but not in women with binge-eating disorder: relationships to co-morbid depression, psychopathology and hormonal variables. *Psychological Medicine, 35*(6), 897-905.

Murphy, R., Nutzinger, D. O., Paul, T., & Leplow, B. (2002). Dissociated conditional-associative learning in anorexia nervosa. *Journal of Clinical and Experimental Neuropsychology, 24*, 176-186.

Nagano-Saito, A., Leyton, M., Monchi, O., Goldberg, Y. K., He, Y., & Dagher, A. (2008). Dopamine depletion impairs frontostriatal functional connectivity during a set-shifting task. *J Neurosci, 28*(14), 3697-3706.

Nakazato, M., Hashimoto, K., Shimizu, E., Kumakiri, C., Koizumi, H., Okamura, N., Mitsumori, M., Komatsu, N., & Iyo, M. (2003). Decreased levels of serum brain-derived neurotrophic factor in female patients with eating disorders. *Biological Psychiatry, 54*(4), 485-490.

Nakazato, M., Hashimoto, K., Yoshimura, K., Hashimoto, T., Shimizu, E., & Iyo, M. (2006). No change between the serum brain-derived neurotrophic factor in female patients with anorexia nervosa before and after partial weight recovery. *Prog Neuropsychopharmacol Biol Psychiatry, 30*(6), 1117-1121.

Nakazato, M., Tchanturia, K., Schmidt, U., Campbell, I. C., Treasure, J., Collier, D. A., Hashimoto, K., & Iyo, M. (2008). Brain-derived neurotrophic factor (BDNF) and set-shifting in currently ill and recovered anorexia nervosa (AN) patients. *Psychological Medicine*, 1-7.

NationalStatistics. (2003). *Better or worse: a longitudinal study of the mental healthy of adults living in private households in Great Britian*. Newport: National Statistics.

Nelson, W. L., Hughes, H. M., Katz, B., & Searight, H. (1999). Anorexic eating attitudes and behaviors of male and female college students. *Adolescence, 34*(135), 621-633.

NICE. (2004). *Eating Disorders: Core interventions in the treatment and magement of anorexia nervosa, bulimia nervosa and related eating disorders.* (Vol. Clinical Guideline 9). London, UK: National Institute for Clinical Excellence.

Nicholls, D., Chater, R., & Lask, B. (2000). Children into DSM don't go: A comparison of classification systems for eating disorders in childhood and early adolescence. *International Journal of Eating Disorders, 28*, 317-324.

Nicolls, D. (2007). Early onset eating disorders in children.   Retrieved 29 June 2008, 2008, from http://bpsu.inopsu.com/studies/eoed/press.html

Ohrmann, P., Kersting, A., Sunslow, T., Lalee-Mentzel, J., Donges, U. S., Feibich, M., Arolt, V., Heindel, W., & Pfleidrer, B. (2004). Proton magnetic resonance spectroscopy in anorexia nervosa: Correlations with cognition. *Neuroreport, 15*, 549-553.

Pendleton-Jones, B., Duncan, C. C., Brouwers, P., & Mirsky, A. F. (1991). Cognition in eating disorders. *Journal of Clinical and Experimental Neuropsychology, 13*, 711-728.

Perianez, J., Rios-Lago, M., Rodriguez-Sanchez, J., Adrover-Roig, D., Sanchez-Cubillo, I., Crespo-Facorro, B., Quemada, J., & Barcelo, F. (2007). Trail Making Test in traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and normative data. *Archives of Clinical Neuropsychology, 22*(4), 433-447.

Powell, T., & Malia, K. (2003). *The Brain Injury Work Book: Exercises for Cognitive Rehabilitation*: Speechmark Publishing Ltd.

Rao, N. P., Reddy, Y. C., Kumar, K. J., Kandavel, T., & Chandrashekar, C. R. (2008). Are neuropsychological deficits trait markers in OCD? *Prog Neuropsychopharmacol Biol Psychiatry, 32*(6), 1574-1579.

Ravizza, S. M., & Carter, C. S. (2008). Shifting set about task switching: behavioral and neural evidence for distinct forms of cognitive flexibility. *Neuropsychologia, 46*(12), 2924-2935.

Redgrave, G. W., Bakker, A., Bello, N. T., Caffo, B. S., Coughlin, J. W., Guarda, A. S., McEntee, J. E., Pekar, J. J., Reinblatt, S. P., Verduzco, G., & Moran, T. H. (2008). Differential brain activation in anorexia nervosa to Fat and Thin words during a Stroop task. *Neuroreport, 19*(12), 1181-1185.

Reichenberg, A., & Harvey, P. D. (2007). Neuropsychological impairments in schizophrenia: Integration of performance-based and brain imaging findings. *Psychological Bulletin, 133*(5), 833-858.

Reitan, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psychology, 19*(5), 393-394.

Reitan, R. M. (1958). Validity of the train making test as indicator of organic brain damage. *Perceptual and Motor Skills, 8*, 271-276.

Reitan, R. M., & Tarshes, E. L. (1959). Differential effects of lateralized brain lesions on the trail making test. *Journal of Nervous and Mental Disease, 129*, 257-262.

Ring, H. A., Baron-Cohen, S., Wheelwright, S., Williams, S. C., Brammer, M., Andrew, C., & Bullmore, E. T. (1999). Cerebral correlates of preserved cognitive skills in autism: a functional MRI study of embedded figures task performance. *Brain, 122 (Pt 7)*, 1305-1315.

Robbins, T. W. (2000). Chemical neuromodulation of frontal-executive functions in humans and other animals. *Experimental Brain Research, 133*(1), 130-138.

Roberts, A. C., De Salvia, M. A., Wilkinson, L. S., Collins, P., Muir, J. L., Everitt, B. J., & Robbins, T. W. (1994). 6-Hydroxydopamine lesions of the prefrontal cortex in monkeys enhance performance on an analog of the Wisconsin Card Sort Test: possible interactions with subcortical dopamine. *Journal of Neuroscience, 14*(5 Pt 1), 2531-2544.

Roberts, M., Dermetriou, L., Treasure, J., & Tchanturia, K. (2007a). Neuropsychological profile in the overweight population: An exploratory

study of set-shifting and detail focussed processing styles. *Therapy, 4*(6), 821-824.

Roberts, M. E. (2006). Disordered eating and obsessive-compulsive symptoms in a sub-clinical student population. *New Zealand Journal of Psychology, 35*(1), 45-54.

Roberts, M. E., Tchanturia, K., Stahl, D., Southgate, L., & Treasure, J. (2007b). A systematic review and meta-analysis of set shifting ability in eating disorders. *Psychological Medicine, 37*(8), 1075-1081.

Roberts, M. E., & Wilson, M. S. (2008). Factor structure and response bias of the Obsessive-Compulsive Inventory-Revised (OCI-R) in a female undergraduate sample from New Zealand. *New Zealand Journal of Psychology, 37*(2), 1-6.

Robin, A. L., Siegel, P. T., Moye, A. W., Gilroy, M., Dennis, A. B., & Sikand, A. (1999). A controlled comparison of family versus individual therapy for adolescents with anorexia nervosa.[see comment]. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(12), 1482-1489.

Robinson, L. J., Thompson, J. M., Gallagher, P., Goswami, U., Young, A. H., Ferrier, I. N., & Moore, P. B. (2006). A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. *Journal of Affective Disorders, 93*, 105-115.

Rosenberg, M. (1965). *Society and the Adolescent Self-image*. Princeton, NJ: Princeton University Press.

Rosenberg, M. (1986). *Conceiving the Self*. Malabar, FL: Krieger.

Rosenberg, M. S., Adams, D. C., & Gurevitch, J. (2000). Metawin: Statistical Software for Meta-analysis (Version 2.0). Sunderland, MA: Sinauer.

Rosval, L., Steiger, H., Bruce, K., Israel, M., Richardson, J., & Aubut, M. (2006). Impulsivity in women with eating disorders: problem of response inhibition, planning, or attention? *International Journal of Eating Disorders, 39*(7), 590-593.

Russell, G. (1979). Bulimia nervosa: An ominous variant of anorexia nervosa. *Psychological Medicine, 9*(3), 429-448.

Russell, G. F., Szmukler, G. I., Dare, C., & Eisler, I. (1987). An evaluation of family therapy in anorexia nervosa and bulimia nervosa. *Archives of General Psychiatry, 44*(12), 1047-1056.

Sachs, G., Schaffer, M., & Winklbaur, B. (2007). Cognitive deficits in bipolar disorder. *Neuropsychiatrie [Article in German], 21*(2), 93-101.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *J Exp Psychol Gen, 132*(4), 566-594.

Santel, S., Baving, L., Krauel, K., Munte, T. F., & Rotte, M. (2006). Hunger and satiety in anorexia nervosa: fMRI during cognitive processing of food pictures. *Brain Res, 1114*(1), 138-148.

Saoud, M., d'Amato, T., Gutknecht, C., Triboulet, P., Bertaud, J. P., Marie-Cardine, M., Dalery, J., & Rochet, T. (2000). Neuropsychological deficit in siblings discordant for schizophrenia. *Schizophrenia Bulletin, 26*(4), 893-902.

Schapman-Williams, A. M., Lock, J., & Couturier, J. (2006). Cognitive-behavioral therapy for adolescents with binge eating syndromes: A case series. *International Journal of Eating Disorders, 39*(3), 252-255.

Schmidt, U. (2005). Epidemiology and aetiology of eating disorders. *Psychiatry, 4*(4), 5-9.

Schmidt, U., Lee, S., Beecham, J., Perkins, S., Treasure, J., Yi, I., Winn, S., Robinson, P., Murphy, R., Keville, S., Johnson-Sabine, E., Jenkins, M., Frost, S., Dodge, L., Berelowits, M., & Eisler, I. (2007). A randomised controlled trial of family therapy and cognitive behaviour therapy guided self-care for adolescents with bulimia nervosa and related disorders. *American Journal of Psychiatry, 164*, 591-598.

Schmidt, U., & Treasure, J. (2006). Anorexia nervosa: valued and visible. A cognitive-interpersonal maintenance model and its implications for research and practice. *British Journal of Clinical Psychology, 45*(3), 343-366.

Sepulveda, A. R., Todd, G., Whitaker, W., Stahl, D., & Treasure, J. (In Progress). An exploratory randomised trial of a support programme for carers of patients with eating disorders.

Serfaty, M., Turkington, D., Heap, M., Ledsham, L., & Jolley, E. (1999). Cognitive therapy versus dietary counselling in the outpatient treatment of anorexia nervosa: effects of the treatment phase. *European Eating Disorders Review, 7*, 334-350.

Shapiro, J. R., Berkman, N. D., Brownley, K. A., Sedway, J. A., Lohr, K. N., & Bulik, C. M. (2007). Bulimia nervosa treatment: A systematic review of randomized controlled trials. *International Journal of Eating Disorders, 40*(4), 321-336.

Sherman, B. J., Savage, C. R., Eddy, K. T., Blais, M. A., Deckersbach, T., Jackson, S. C., Franko, D. L., Rauch, S. L., & Herzog, D. B. (2006). Strategic memory in adults with anorexia nervosa: are there similarities to obsessive compulsive spectrum disorders? *International Journal of Eating Disorders, 39*(6), 468-476.

Shin, M. S., Choi, H., Kim, H., Hwang, J. W., Kim, B. N., & Cho, S. C. (2008). A study of neuropsychological deficit in children with obsessive-compulsive disorder. *European Psychiatry*.

Sitskoorn, M. M., Aleman, A., Ebisch, S. J., Appels, M. C., & Kahn, R. S. (2004). Cognitive deficits in relatives of patients with schizophrenia: a meta-analysis. *Schizophrenia Research, 71*(2-3), 285-295.

Snitz, B. E., Macdonald III, A. W., & Carter, C. S. (2006). Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of puntative endophenotypes. *Schizophrenia Bulletin, 32*, 179-194.

Snodgrass, J. G., Smith, B., Feenan, K., & Corwin, J. (1987). Fragmenting pictures on the Apple Macintosh computer for experimental and clinical applications. *Behavior Research Methods, Instruments & Computers, 19*(2), 270-274.

Southgate, L. (2005). *Response inhibition in anorexia nervosa and bulimia nervosa: an exploration of neuropsychological functions and their association with personality traits and behaviours.* King's College University of London, London.

Southgate, L., Tchanturia, K., Collier, D., & Treasure, J. (2008a). The development of the childhood retrospective perfectionism questionnaire (CHIRP) in an eating disorder sample. *European Eating Disorders Review*.

Southgate, L., Tchanturia, K., & Treasure, J. (2007). Information processing bias in anorexia nervosa. *Psychiatry Research, doi 10.1016j psycres. 2007.07.017*.

Southgate, L., Tchanturia, K., & Treasure, J. (2008b). Information processing bias in anorexia nervosa. *Psychiatry Research, 160*(2), 221-227.

Spencer, J., O'Brien, J., Riggs, K., Braddick, O., Atkinson, J., & Wattam-Bell, J. (2000). Motion processing in autism: evidence for a dorsal stream deficiency. *Neuroreport, 11*(12), 2765-2767.

Speranza, M., Corcos, M., Godart, N., Loas, G., Guilbaud, O., Jeammet, P., & Flament, M. (2001). Obsessive compulsive disorders in eating disorders. *Eating Behaviors, 2*(3), 193-207.

Spreen, O., & Strauss, E. (1998). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary* (2nd ed.). New York: Oxford University Press.

Steichen, T. (1998). sbe19: tests for publication bias in meta-analysis. *Stata Technical Bulletin, 41*, 9-15.

Steiger, H., & Bruce, K. R. (2007). Phenotypes, endophenotypes, and genotypes in bulimia spectrum eating disorders. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie, 52*(4), 220-227.

Steiger, H., Gauvin, L., Joober, R., Israel, M., Ng Ying Kin, N. M., Bruce, K. R., Richardson, J., Young, S. N., & Hakim, J. (2006). Intrafamilial correspondences on platelet [3H-]paroxetine-binding indices in bulimic probands and their unaffected first-degree relatives. *Neuropsychopharmacology, 31*(8), 1785-1792.

Steinglass, J. E., Walsh, B., & Stern, Y. (2006). Set shifting deficit in anorexia nervosa. *Journal of the International Neuropsychological Society, 12*(3), 431-435.

Steinhausen, H. C. (2002). The outcome of anorexia nervosa in the 20th century. *American Journal of Psychiatry, 159*(8), 1284-1293.

Steketee, G., Frost, R., & Bogart, K. (1996). The Yale-Brown obsessive compulsive scale: Interview versus self-report. *Behaviour Research and Therapy, 34*(8), 675-684.

Stice, E. (2004). Risk and maintenance factors for eating pathology: A meta-analytic review. *Psychological Bulletin, 128*, 825-848.

Stober, J. (1998). The Frost Multidimensional Perfectionism Scale revisited: More perfect with four (instead of six) dimensions. *Personality and Individual Differences, 24*(4), 481-491.

Sunday, S. R., & Halmi, K. A. (2000). Comparison of the yale-brown-cornell eating disorders scale in recovered eating disorder patients, restrained dieters, and nondieting controls. *International Journal of Eating Disorders, 28*(4), 455-459.

Tchanturia, K., Anderluh, M. B., Morris, R. G., Rabe-Hesketh, S., Collier, D. A., Sanchez, P., & Treasure, J. L. (2004a). Cognitive flexibility in anorexia nervosa and bulimia nervosa. *Journal of the International Neuropsychological Society, 10*(4), 513-520.

Tchanturia, K., Campbell, l. C., Morris, R., & Treasure, J. (2005). Neuropsychological Studies in Anorexia Nervosa. *International Journal of Eating Disorders, 37*(Suppl), S72-S76.

Tchanturia, K., Davies, H., & Campbell, I. C. (2007). Cognitive Remediation for patients with Anorexia Nervosa: preliminary findings. *Annals of General Psychiatry, 14*, 1-6.

Tchanturia, K., Davies, H., Lopez, C., Schmidt, U., Treasure, J., & Wykes, T. (2008). Neuropsychological task performance before and after cognitive remediation in anorexia nervosa: a pilot case-series. *Psychological Medicine, 38*(9), 1371-1373.

Tchanturia, K., Morris, R., Anderluh, M., Collier, D., Nikolaou, V., & Treasure, J. (2004b). Set shifting in anorexia nervosa: An examination before and after weight gain, in full recovery and relationship to childhood and adult OCPD traits. *Journal of Psychiatric Research, 38*(5), 545-552.

Tchanturia, K., Morris, R., Surguladze, S., & Treasure, J. (2002). An examination of perceptual and cognitive set shifting tasks in acute anorexic nervosa and following recovery. *Eating and Weight Disorders, 7*(4), 312-315.

Tchanturia, K., Serpell, L., Troop, N., & Treasure, J. (2001). Perceptual illusions in eating disorders: Rigid and fluctuating styles. *Journal of Behavior Therapy and Experimental Psychiatry, 32*(3), 107-115.

Thompson, S. N. B. (1993). Implications of neuropsychological test results of women in a new phase of anorexia nervosa. *European Eating Disorders Review, 1*, 152-165.

Tokley, M., & Kemps, E. (2007). Preoccupation with detail contributes to poor abstraction in women with anorexia nervosa. *Journal of Clinical and Experimental Neuropsychology, 29*, 734-741.

Touyz, S. W., Beumont, P. J. V., & Johnstone, L. C. (1986). Neuropsychological correlates of dieting disorders. *International Journal of Eating Disorders, 5*, 1025-1034.

Treasure, J., Lopez, C., & Roberts, M. E. (2007). Endophenotypes in eating disorders: moving toward etiologically based diagnosis and treatment focussed on pathophysiology. *Pediatric Health, 1*(2), 171-181.

Treasure, J., Murphy, T., Szmukler, T., Todd, G., Gavan, K., & Joyce, J. (2001). The experience of caregiving for severe mental illness: A comparison between anorexia nervosa and psychosis. *Social Psychiatry and Psychiatric Epidemiology, 36*(7), 343-347.

Treasure, J. L. (2007). Getting beneath the phenotype of anorexia nervosa: the search for viable endophenotypes and genotypes. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie, 52*(4), 212-219.

Treasure, J. L., Wack, E. R., & Roberts, M. E. (2008). Models as a high risk group: the health implications of a size zero culture. *British Journal of Psychiatry, 192*, 243-244.

Uznadze, D. N. (1966). *The Psycholoogy of Set*. New York: Consultants' Bureau.

van Son, G. E., van Hoeken, D., Bartelds, A. I., van Furth, E. F., & Hoek, H. W. (2006). Urbanisation and the incidence of eating disorders. *British Journal of Psychiatry, 189*(6), 562-563.

Wade, T. D., Bergin, J. L., Tiggemann, M., Bulik, C. M., & Fairburn, C. G. (2006). Prevalence and long-term course of lifetime eating disorders in an adult Australian twin cohort. *Australian and New Zealand Journal of Psychiatry, 40*(2), 121-128.

Wagner, A., Aizenstein, H., Mazurkewicz, L., Fudge, J., Frank, G. K., Putnam, K., Bailer, U. F., Fischer, L., & Kaye, W. H. (2008). Altered insula response to taste stimuli in individuals recovered from restricting-type anorexia nervosa. *Neuropsychopharmacology, 33*(3), 513-523.

Wagner, A., Aizenstein, H., Venkatraman, V. K., Fudge, J., May, J. C., Mazurkewicz, L., Frank, G. K., Bailer, U. F., Fischer, L., Nguyen, V., Carter, C., Putnam, K., & Kaye, W. H. (2007). Altered reward processing in women recovered from anorexia nervosa. *Am J Psychiatry, 164*(12), 1842-1849.

Waiter, G. D., Williams, J. H., Murray, A. D., Gilchrist, A., Perrett, D. I., & Whiten, A. (2005). Structural white matter deficits in high-functioning individuals

with autistic spectrum disorder: a voxel-based investigation. *Neuroimage, 24*(2), 455-461.

Wallin, U., Kronovall, P., & Majewski, M.-L. (2000). Body awareness therapy in teenage anorexia nervosa: outcome after 2 years. *European Eating Disorders Review, 8*, 19-30.

Walters, J. T. R., & Owen, M. J. (2007). Endophenotypes in psychiatric genetics. *Molecular Psychiatry, 12*, 886-890.

Wentz, E., Gillberg, C., Gillberg, I. C., & Rastam, M. (2001). Ten year follow-up of adolescent-onset anorexia nervosa: psychoatric disorders and overall functioning scales. *Journal of Child Psychology and Psychiatry, 42*(5), 613-622.

Whitney, J., Murray, J., Gavan, K., Todd, G., Whitaker, W., & Treasure, J. (2005). Experience of caring for someone with anorexia nervosa: Qualitative study. *British Journal of Psychiatry, 187*(5), 444-449.

Wilksch, S. M., & Wade, T. D. (2008). An investigation of temperament endophenotype candidates for early emergence of the core cognitive component of eating disorders. *Psychol Med*, 1-11.

Wilsdon, A., & Wade, T. D. (2006). Executive functioning in anorexia nervosa: exploration of the role of obsessionality, depression and starvation. *Journal of Psychiatric Research, 40*(8), 746-754.

Witkin, H. A. (1971). *A Manual for the Embedded Figure Test*. California: Consulting Psychologists Press.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (2002). *Group Embedded Figures Test: Manual*: Mind Garden, Inc.

Witt, E. D., Ryan, C., & George Hsu, L. K. (1985). Learning deficits in adolescents with anorexia nervosa. *Journal of Nervous and Mental Disease, 173*, 182-184.

Wobrock, T., Ecker, U. K., Scherk, H., Schneider-Axmann, T., Falkai, P., & Gruber, O. (2008). Cognitive impairment of executive function as a core symptom of schizophrenia. *World J Biol Psychiatry*, 1-10.

Yurgelun-Todd, D. A., & Kinney, D. K. (1993). Patterns of neuropsychological deficits that discriminate schizophrenic individuals from siblings and control subjects. *Journal of Neuropsychiatry and Clinical Neurosciences, 5*(3), 294-300.

Zastrow, A., Kaiser, S., Stippich, C., Walther, S., Herzog, W., Tchanturia, K., Belger, A., Weisbrod, M., Treasure, J., & Friederich, H. C. (2009). Neural Correlates of Impaired Cognitive-Behavioral Flexibility in Anorexia Nervosa. *Am J Psychiatry*.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*(6), 361-370.